

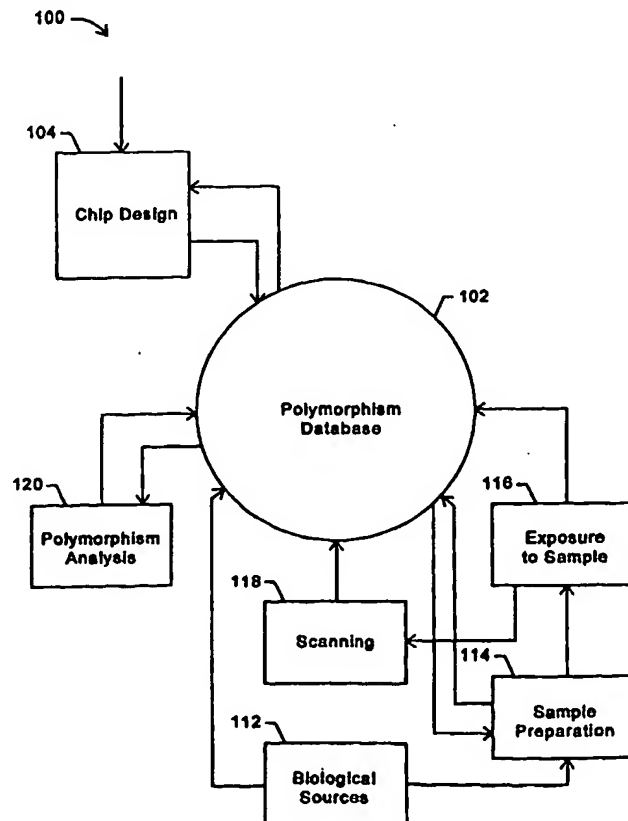


## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

|   |           |  |
|---|-----------|--|
| <b>(51) International Patent Classification <sup>6</sup> :</b><br><b>C12Q 1/68, G06F 17/30 // 159:00</b>  | <b>A1</b> | <b>(11) International Publication Number:</b> <b>WO 99/05324</b><br><b>(43) International Publication Date:</b> 4 February 1999 (04.02.99)   |
| <b>(21) International Application Number:</b> PCT/US98/15458<br><b>(22) International Filing Date:</b> 24 July 1998 (24.07.98)<br><br><b>(30) Priority Data:</b><br>60/053,842      25 July 1997 (25.07.97)      US<br>60/069,198      11 December 1997 (11.12.97)      US<br>60/069,436      11 December 1997 (11.12.97)      US<br><br><b>(71) Applicant:</b> AFFYMETRIX, INC. [US/US]; 3380 Central Expressway, Santa Clara, CA 95051 (US).<br><br><b>(72) Inventors:</b> BALABAN, David, J.; 37 Bret Harte Road, San Rafael, CA 94901 (US). BAID, Joyti; 3330 Estates Drive, San Jose, CA 95148 (US). BERNIO, Anthony; 570 South 12th Street, San Jose, CA 95112 (US).<br><br><b>(74) Agents:</b> LANG, Dan, H. et al.; Townsend and Townsend and Crew LLP, 8th floor, Two Embarcadero Center, San Francisco, CA 94111-3834 (US). |           | <b>(81) Designated States:</b> JP, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).<br><br><b>Published</b><br><i>With international search report.</i> |

**(54) Title:** SYSTEM FOR PROVIDING A POLYMORPHISM DATABASE**(57) Abstract**

Systems and methods for organising information relating to a study of polymorphisms. A database model (102) is provided which interrelates information about one or more of, e.g. subjects (112) from whom samples (114) are extracted, primers used in extracting the DNA from the subjects, about the samples themselves, about experiments done on samples, about particular oligonucleotide probe arrays used to perform experiments, about analysis procedures performed on the samples, and about analysis results. The model is readily translatable into database languages such as SQL. The database model scales to permit storage of information about large numbers of subjects, samples, experiments, chips, etc.



**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

|    |                          |    |  |    |  |    |                          |
|----|--------------------------|----|--|----|--|----|--------------------------|
| AL | Albania                  | ES | Spain                                    | LS | Lesotho                                      | SI | Slovenia                 |
| AM | Armenia                  | FI | Finland                                  | LT | Lithuania                                    | SK | Slovakia                 |
| AT | Austria                  | FR | France                                   | LU | Luxembourg                                   | SN | Senegal                  |
| AU | Australia                | GA | Gabon                                    | LV | Latvia                                       | SZ | Swaziland                |
| AZ | Azerbaijan               | GB | United Kingdom                           | MC | Monaco                                       | TD | Chad                     |
| BA | Bosnia and Herzegovina   | GE | Georgia                                  | MD | Republic of Moldova                          | TG | Togo                     |
| BB | Barbados                 | GH | Ghana                                    | MG | Madagascar                                   | TJ | Tajikistan               |
| BE | Belgium                  | GN | Guinea                                   | MK | The former Yugoslav<br>Republic of Macedonia | TM | Turkmenistan             |
| BF | Burkina Faso             | GR | Greece                                   | ML | Mali   | TR | Turkey                   |
| BG | Bulgaria                 | HU | Hungary                                  | MN | Mongolia                                     | TT | Trinidad and Tobago      |
| BJ | Benin                    | IE | Ireland                                  | MR | Mauritania                                   | UA | Ukraine                  |
| BR | Brazil                   | IL | Israel                                   | MW | Malawi                                       | UG | Uganda                   |
| BY | Belarus                  | IS | Iceland                                  | MX | Mexico                                       | US | United States of America |
| CA | Canada                   | IT | Italy                                    | NE | Niger  | UZ | Uzbekistan               |
| CF | Central African Republic | JP | Japan                                    | NL | Netherlands                                  | VN | Viet Nam                 |
| CG | Congo                    | KE | Kenya                                    | NO | Norway                                       | YU | Yugoslavia               |
| CH | Switzerland              | KG | Kyrgyzstan                               | NZ | New Zealand                                  | ZW | Zimbabwe                 |
| CI | Côte d'Ivoire            | KP | Democratic People's<br>Republic of Korea | PL | Poland                                       |    |                          |
| CM | Cameroon                 | KR | Republic of Korea                        | PT | Portugal                                     |    |                          |
| CN | China                    | KZ | Kazakhstan                               | RO | Romania                                      |    |                          |
| CU | Cuba                     | LC | Saint Lucia                              | RU | Russian Federation                           |    |                          |
| CZ | Czech Republic           | LI | Liechtenstein                            | SD | Sudan  |    |                          |
| DE | Germany                  | LK | Sri Lanka                                | SE | Sweden                                       |    |                          |
| DK | Denmark                  | LR | Liberia                                  | SG | Singapore                                    |    |                          |
| EE | Estonia                  |    |  |    |  |    |                          |

## SYSTEM FOR PROVIDING A POLYMORPHISM DATABASE

### CROSS-REFERENCE TO RELATED APPLICATIONS

The present application claims priority from U.S. Prov. App. No. 60/053,842 filed July 25, 1997, entitled COMPREHENSIVE BIO-INFORMATICS DATABASE, from U.S. Prov. App. No. 60/069,198 filed on December 11, 1997, entitled COMPREHENSIVE DATABASE FOR BIOINFORMATICS, and from U.S. Prov. App. No. 60/069,436, entitled GENE EXPRESSION AND EVALUATION SYSTEM, filed on December 11, 1997. The contents of all three provisional applications are herein incorporated by reference.

The subject matter of the present application is related to the subject matter of the following three co-assigned applications filed on the same day as the present application. GENE EXPRESSION AND EVALUATION SYSTEM (Attorney Docket No. 018547-035010), METHOD AND APPARATUS FOR PROVIDING A BIOINFORMATICS DATABASE (Attorney Docket No. 018547-033810), METHOD AND SYSTEM FOR PROVIDING A PROBE ARRAY CHIP DESIGN DATABASE (Attorney Docket No. 018547-033830). The contents of these three applications are herein incorporated by reference.

### BACKGROUND OF THE INVENTION

The present invention relates to the collection and storage of information pertaining to chips for processing biological samples and thereby identifying polymorphisms.

The genomes of all organisms undergo spontaneous mutation in the course of their continuing evolution generating variant forms of progenitor sequences (Gusella, *Ann. Rev. Biochem.* 55, 831-854 (1986)). The variant form may confer an evolutionary advantage or disadvantage relative to a progenitor form or may be neutral. In some instances, a variant form confers a lethal disadvantage and is not transmitted to subsequent generations of the organism. In other instances, a variant form confers an

evolutionary advantage to the species and is eventually incorporated into the DNA of many or most members of the species and effectively becomes the progenitor form. In many instances, both progenitor and variant form(s) survive and co-exist in a species population. The coexistence of multiple forms of a sequence gives rise to polymorphisms.

Despite the increased amount of nucleotide sequence data being generated in recent years, only a minute proportion of the total repository of polymorphisms in humans and other organisms has so far been identified. The paucity of polymorphisms hitherto identified is due to the large amount of work required for their detection by conventional methods. For example, a conventional approach to identifying polymorphisms might be to sequence the same stretch of oligonucleotides in a population of individuals by dideoxy sequencing. In this type of approach, the amount of work increases in proportion to both the length of sequence and the number of individuals in a population and becomes impractical for large stretches of DNA or large numbers of persons.

Devices and computer systems for forming and using arrays of materials on a substrate have been developed. These devices and systems have been used for identifying polymorphisms. For example, PCT application WO92/10588, incorporated herein by reference for all purposes, describes techniques for sequencing or sequence checking nucleic acids and other materials. Arrays for performing these operations may be formed in arrays according to the methods of, for example, the pioneering techniques disclosed in U.S. Patent No. 5,143,854 and U.S. Patent No. 5,571,639, both incorporated herein by reference for all purposes.

According to one aspect of the techniques described therein, an array of nucleic acid probes is fabricated at known locations on a chip or substrate. A fluorescently labeled nucleic acid is then brought into contact with the chip and a scanner generates an image file indicating the locations where the labeled nucleic acids bound to the chip. Based upon the identities of the probes at these locations, it becomes possible to extract information such as the identity of polymorphic forms in of DNA or RNA. Such systems have been used to form, for example, arrays of DNA that may be used to study and detect mutations relevant to cystic fibrosis, the P53 gene (relevant to certain cancers), HIV, and other genetic characteristics.

It would be highly useful to apply such arrays to the study of polymorphisms on a large scale. For example, it would be useful to conduct large scale studies on the correlation between certain polymorphisms and individual characteristics such as susceptibility to diseases and effectiveness of drug treatments. To achieve these benefits, it is contemplated that the operations of chip design, construction, sample preparation, and analysis will occur on a very large scale. The quantity of information related to each of these steps to store and correlate is vast. For large scale polymorphism studies, it will be necessary to store this information in a way to facilitate later advantageous querying and retrieval. What is needed is a system and method suitable for storing and organizing large quantities of information used in conjunction with polymorphism studies.

### SUMMARY OF THE INVENTION

The present invention provides systems and methods for organizing information relating to study of polymorphisms. A database model is provided which interrelates information about one or more of, e.g, subjects from whom samples are extracted, primers used in extracting the DNA from the subjects, about the samples themselves, about experiments done on samples, about particular oligonucleotide probe arrays used to perform experiments, about analysis procedures performed on the samples, and about analysis results. The model is readily translatable into database languages such as SQL. The database model scales to permit storage of information about large numbers of subjects, samples, experiments, chips, etc.

Applications include linkage studies to determine resistance to drugs, susceptibility to diseases, and study of every characteristic of humans and other organisms that is related genetic variability. Another application of a database constructed according to this model is quality control of the various steps of performing a polymorphism study. By preserving information about every step of a polymorphism study, one can assess the reliability of the results or use the preserved information as feedback to improve procedures.

A further understanding of the nature and advantages of the inventions herein may be realized by reference to the remaining portions of the specification and the attached drawings.

## BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 illustrates an overall system and process for forming and analyzing arrays of biological materials such as DNA or RNA.

Fig. 2A illustrates a computer system suitable for use in conjunction with the overall system of Fig. 1.

Fig. 2B illustrates a computer network suitable for use in conjunction with the overall system of Fig. 1.

Fig. 3 illustrates a key for interpreting a database model.

Figs. 4A-4H illustrate a database model for maintaining information for the system and process of Fig. 1 according to one embodiment of the present invention.

## DESCRIPTION OF SPECIFIC EMBODIMENTS

### Investigation of Polymorphisms

#### A. Preparation of Samples

Polymorphisms are detected in a target nucleic acid from an individual being analyzed. For assay of genomic DNA, virtually any biological sample (other than pure red blood cells) is suitable. For example, convenient tissue samples include whole blood, semen, saliva, tears, urine, fecal material, sweat, buccal, skin and hair. For assay of cDNA or mRNA, the tissue sample must be obtained from an organ in which the target nucleic acid is expressed. For example, if the target nucleic acid is a cytochrome P450, the liver is a suitable source.

Many of the methods described below require amplification of DNA from target samples. This can be accomplished by e.g., PCR. *See generally PCR Technology: Principles and Applications for DNA Amplification* (ed. H.A. Erlich, Freeman Press, NY, NY, 1992); *PCR Protocols: A Guide to Methods and Applications* (eds. Innis, et al., Academic Press, San Diego, CA, 1990); Mattila et al., *Nucleic Acids Res.* 19, 4967 (1991); Eckert et al., *PCR Methods and Applications* 1, 17 (1991); *PCR* (eds. McPherson et al., IRL Press, Oxford); and U.S. Patent 4,683,202 (each of which is incorporated by reference for all purposes).

Other suitable amplification methods include the ligase chain reaction (LCR) (see Wu and Wallace, *Genomics* 4, 560 (1989), Landegren et al., *Science* 241, 1077 (1988), transcription amplification (Kwoh et al., *Proc. Natl. Acad. Sci. USA* 86, 1173 (1989)), and self-sustained sequence replication (Guatelli et al., *Proc. Nat. Acad.*

Sci. USA, 87, 1874 (1990)) and nucleic acid based sequence amplification (NASBA). The latter two amplification methods involve isothermal reactions based on isothermal transcription, which produce both single stranded RNA (ssRNA) and double stranded DNA (dsDNA) as the amplification products in a ratio of about 30 or 100 to 1, respectively.

#### B. Detection of Polymorphisms in Target DNA

There are two distinct types of analysis depending whether a polymorphism in question has already been characterized. The first type of analysis is sometimes referred to as de novo characterization. This analysis compares target sequences in different individuals to identify points of variation, i.e., polymorphic sites. By analyzing groups of individuals representing the greatest ethnic diversity among humans and greatest breed and species variety in plants and animals, patterns characteristic of the most common alleles/haplotypes of the locus can be identified, and the frequencies of such populations in the population determined. Additional allelic frequencies can be determined for subpopulations characterized by criteria such as geography, race, or gender. The second type of analysis is determining which form(s) of a characterized polymorphism are present in individuals under test. There are a variety of suitable procedures, which are discussed in turn.

##### 1. Allele-Specific Probes

The design and use of allele-specific probes for analyzing polymorphisms is described by e.g., Saiki et al., *Nature* 324, 163-166 (1986); Dattagupta, EP 235,726, Saiki, WO 89/11548. Allele-specific probes can be designed that hybridize to a segment of target DNA from one individual but do not hybridize to the corresponding segment from another individual due to the presence of different polymorphic forms in the respective segments from the two individuals. Hybridization conditions should be sufficiently stringent that there is a significant difference in hybridization intensity between alleles, and preferably an essentially binary response, whereby a probe hybridizes to only one of the alleles. Some probes are designed to hybridize to a segment of target DNA such that the polymorphic site aligns with a central position (e.g., in a 15 mer at the 7 position; in a 16 mer, at either the 8 or 9 position) of the

probe. This design of probe achieves good discrimination in hybridization between different allelic forms.

Allele-specific probes are often used in pairs, one member of a pair showing a perfect match to a reference form of a target sequence and the other member showing a perfect match to a variant form. Several pairs of probes can then be immobilized on the same support for simultaneous analysis of multiple polymorphisms within the same target sequence.

## 2. Tiling Arrays

The polymorphisms can also be identified by hybridization to nucleic acid arrays, some example of which are described by WO 95/11995 (incorporated by reference in its entirety for all purposes). WO 95/11995 also describes subarrays that are optimized for detection of a variant forms of a precharacterized polymorphism. Such a subarray contains probes designed to be complementary to a second reference sequence, which is an allelic variant of the first reference sequence. The second group of probes is designed by the same principles as described in the Examples except that the probes exhibit complementarity to the second reference sequence. The inclusion of a second group (or further groups) can be particular useful for analyzing short subsequences of the primary reference sequence in which multiple mutations are expected to occur within a short distance commensurate with the length of the probes (*i.e.*, two or more mutations within 9 to 21 bases).

## 3. Allele-Specific Primers

An allele-specific primer hybridizes to a site on target DNA overlapping a polymorphism and only primes amplification of an allelic form to which the primer exhibits perfect complementarity. See Gibbs, *Nucleic Acid Res.* 17, 2427-2448 (1989).

This primer is used in conjunction with a second primer which hybridizes at a distal site. Amplification proceeds from the two primers leading to a detectable product signifying the particular allelic form is present. A control is usually performed with a second pair of primers, one of which shows a single base mismatch at the polymorphic site and the other of which exhibits perfect complementarity to a distal site. The single-base mismatch prevents amplification and no detectable product is formed. The method works best when the mismatch is included in the 3'-most position of the oligonucleotide aligned with the polymorphism because this position is most destabilizing to elongation from the primer. See, *e.g.*, WO 93/22456.



#### 4. Direct-Sequencing

The direct analysis of the sequence of polymorphisms of the present invention can be accomplished using either the dideoxy chain termination method or the Maxam Gilbert method (see Sambrook et al., *Molecular Cloning, A Laboratory Manual* (2nd Ed., CSHP, New York 1989); Zyskind et al., *Recombinant DNA Laboratory Manual*, (Acad. Press, 1988)).

#### 5. Denaturing Gradient Gel Electrophoresis

Amplification products generated using the polymerase chain reaction can be analyzed by the use of denaturing gradient gel electrophoresis. Different alleles can be identified based on the different sequence-dependent melting properties and electrophoretic migration of DNA in solution. Erlich, ed., *PCR Technology, Principles and Applications for DNA Amplification*, (W.H. Freeman and Co, New York, 1992), Chapter 7.

#### 6. Single-Strand Conformation Polymorphism Analysis

Alleles of target sequences can be differentiated using single-strand conformation polymorphism analysis, which identifies base differences by alteration in electrophoretic migration of single stranded PCR products, as described in Orita et al., *Proc. Nat. Acad. Sci.* 86, 2766-2770 (1989). Amplified PCR products can be generated as described above, and heated or otherwise denatured, to form single stranded amplification products. Single-stranded nucleic acids may refold or form secondary structures which are partially dependent on the base sequence. The different electrophoretic mobilities of single-stranded amplification products can be related to base-sequence difference between alleles of target sequences.

#### Biological Material Analysis System

One embodiment of the present invention operates in the context of a system for analyzing biological or other materials using arrays that themselves include probes that may be made of biological materials such as RNA or DNA. The VLSIPS™ and GeneChip™ technologies provide methods of making and using very large arrays of polymers, such as nucleic acids, on chips. See U.S. Patent No. 5,143,854 and PCT Patent Publication Nos. WO 90/15070 and 92/10092, each of which is hereby incorporated by reference for all purposes. Nucleic acid probes on the chip are used to

detect complementary nucleic acid sequences in a sample nucleic acid of interest (the "target" nucleic acid).

Fig. 1 illustrates an overall system 100 for forming and analyzing arrays of biological materials such as RNA or DNA. A part of system 100 is a polymorphism database 102. Polymorphism database 102 includes information about, e.g., biological sources, preparation of samples, design of arrays, raw data obtained from applying experiments to chips, analysis procedures applied, and analysis results, etc. Polymorphism database 102 facilitates large scale study of polymorphisms.

A chip design system 104 is used to design arrays of polymers such as biological polymers such as RNA or DNA. Chip design system 104 may be, for example, an appropriately programmed Sun Workstation or personal computer or workstation, such as an IBM PC equivalent, including appropriate memory and a CPU. Chip design system 104 obtains inputs from a user regarding chip design objectives including polymorphisms of interest, and other inputs regarding the desired features of the array. Optionally, chip design system 104 obtains inputs from external databases such as GenBank. The output of chip design system 104 is a set of chip design computer files in the form of, for example, a switch matrix, as described in PCT application WO 92/10092, and other associated computer files. The chip design computer files form a part of polymorphism database 102. Systems for designing chips for study of polymorphisms are disclosed in U.S. Patent No. 5,571,639 and in PCT application WO 95/11995, the contents of which are herein incorporated by reference.

The chip design files are input to a mask design system (not shown) that designs the lithographic masks used in the fabrication of arrays of molecules such as DNA. The mask design system designs the lithographic masks used in the fabrication of probe arrays. The mask design system generates mask design files that are then used by a mask construction system (not shown) to construct masks or other synthesis patterns such as chrome-on-glass masks for use in the fabrication of polymer arrays.

The masks are used in a synthesis system (not shown). The synthesis system includes the necessary hardware and software used to fabricate arrays of polymers on a substrate or chip. The synthesis system includes a light source and a chemical flow cell on which the substrate or chip is placed. A mask is placed between the light source and the substrate/chip, and the two are translated relative to each other at appropriate times for deprotection of selected regions of the chip. Selected chemical reagents are

directed through the flow cell for coupling to deprotected regions, as well as for washing and other operations. The substrates fabricated by the synthesis system are optionally diced into smaller chips. The output of the synthesis system is a chip ready for application of a target sample.

5                   Information about the mask design, mask construction, and probe array synthesis is presented by way of background. A biological source 112 is, for example, tissue from a plant or animal. Various processing steps are applied to material from biological source 112 by a sample preparation system 114. Operation of sample preparation system 114 in the context of a polymorphism study is discussed below in  
10 further detail.

                  The prepared samples include nucleic acid sequences such as DNA. When the sample is applied to the chip by a sample exposure system 116, the nucleic acids may or may not bond to the probes. The nucleic acids can be tagged with fluorescein labels to determine which probes have bonded to nucleotide sequences from  
15 the sample. The prepared samples will be placed in a scanning system 118. Scanning system 118 includes a detection device such as a confocal microscope or CCD (charge-coupled device) that is used to detect the location where labeled receptors have bound to the substrate. The output of scanning system 118 is an image file(s) indicating, in the case of fluorescein labeled receptor, the fluorescence intensity (photon counts or other  
20 related measurements, such as voltage) as a function of position on the substrate. These image files may also form a part of polymorphism database 102. Since higher photon counts will be observed where the labeled nucleic acid(s) has bound more strongly to the array of probes, and since the monomer sequence of the probes on the substrate is known as a function of position, it becomes possible to analyze the sequence(s) of the nucleic  
25 acid(s) that are complementary to the probes.

                  The image files and the design of the chips are input to an analysis system 120 that, e.g., calls bases. Such analysis techniques are described in EPO Pub. No. 0717113A, the contents of which are herein incorporated by reference.

                  Chip design system 104, analysis system 120 and control portions of  
30 exposure system 116, sample preparation system 114, and scanning system 118 may be appropriately programmed computers such as a Sun workstation or IBM-compatible PC. An independent computer for each system may perform the computer-implemented functions of these systems or one computer may combine the computerized functions of

two or more systems. One or more computers may maintain chip design database 102 independent of the computers operating the systems of Fig. 1 or chip design database 102 may be fully or partially maintained by these computers.

Fig. 2A depicts a block diagram of a host computer system 10 suitable for implementing the present invention. Host computer system 210 includes a bus 212 which interconnects major subsystems such as a central processor 214, a system memory 216 (typically RAM), an input/output (I/O) adapter 218, an external device such as a display screen 224 via a display adapter 226, a keyboard 232 and a mouse 234 via an I/O adapter 218, a SCSI host adapter 236, and a floppy disk drive 238 operative to receive a floppy disk 240. SCSI host adapter 236 may act as a storage interface to a fixed disk drive 242 or a CD-ROM player 244 operative to receive a CD-ROM 246. Fixed disk 244 may be a part of host computer system 210 or may be separate and accessed through other interface systems. A network interface 248 may provide a direct connection to a remote server via a telephone link or to the Internet. Network interface 248 may also connect to a local area network (LAN) or other network interconnecting many computer systems. Many other devices or subsystems (not shown) may be connected in a similar manner.

Also, it is not necessary for all of the devices shown in Fig. 2A to be present to practice the present invention, as discussed below. The devices and subsystems may be interconnected in different ways from that shown in Fig. 2A. The operation of a computer system such as that shown in Fig. 2A is readily known in the art and is not discussed in detail in this application. Code to implement the present invention, may be operably disposed or stored in computer-readable storage media such as system memory 216, fixed disk 242, CD-ROM 246, or floppy disk 240.

Fig. 2B depicts a network 260 interconnecting multiple computer systems 210. Network 260 may be a local area network (LAN), wide area network (WAN), etc. Bioinformatics database 102 and the computer-related operations of the other elements of Fig. 2B may be divided amongst computer systems 210 in any way with network 260 being used to communicate information among the various computers. Portable storage media such as floppy disks may be used to carry information between computers instead of network 260.

Overall Description of Database

Polymorphism database 102 is preferably a relational database with a complex internal structure. The structure and contents of chip design database 102 will be described with reference to a logical model depicted in Figs. 4A-4H that describes the contents of tables of the database as well as interrelationships among the tables. A visual depiction of this model will be an Entity Relationship Diagram (ERD) which includes entities, relationships, and attributes. A detailed discussion of ERDs is found in "ERwin version 3.0 Methods Guide" available from Logic Works, Inc. of Princeton, NJ, the contents of which are herein incorporated by reference. Those of skill in the art will appreciate that automated tools such as Developer 2000 available from Oracle will convert the ERD from Figs. 4A-4H directly into executable code such as SQL code for creating and operating the database.

Fig. 3 is a key to the ERD that will be used to describe the contents of chip design database 102. A representative table 302 includes one or more key attributes 304 and one or more non-key attributes 306. Representative table 302 includes one or more records where each record includes fields corresponding to the listed attributes. The contents of the key fields taken together identify an individual record. In the ERD, each table is represented by a rectangle divided by a horizontal line. The fields or attributes above the line are key while the fields or attributes below the line are non-key. An identifying relationship 308 signifies that the key attribute of a parent table 310 is also a key attribute of a child table 312. A non-identifying relationship 314 signifies that the key attribute of a parent table 316 is also a non-key attribute of a child table 318. Where (FK) appears in parenthesis, it indicates that an attribute of one table is a key attribute of another table. Both the depicted non-identifying and identifying relationship are one to one-or-more relationships where one record in the parent table corresponds to one or more records in the child table. An alternative non-identifying relationship 324 is a one to zero-or-more relationship where one record in a parent table 320 corresponds to zero or more records in a child table 322.

Database Model

Figs. 4A-4H are entity relationship diagrams (ERDs) showing elements of polymorphism database 102 according to one embodiment of the present invention. Each

rectangle in the diagram corresponds to a table in database 102. First, the relationships and general contents of the various tables will be described.

The interrelationships and general contents of the tables of database 102 will be described first. Then a chart will be presented listing and describing all of the fields of the various tables.

Fig. 4A illustrates core elements of database 102 according to one embodiment of the present invention. A subject table 402 lists organisms from which samples have been extracted for polymorphism analysis or other tissue sources. Samples may also be obtained from tissue collections not associated with any one identified organism. Information stored within subject table 402 includes the name, gender, family, position with family, (e.g., father, mother, etc.), and ethnic group. For human subjects, the name and family will preferably be represented in coded form to assure privacy. Associated with each subject is a species as listed in a species table 404. Also, a relationship may be defined among subjects a subject relationship table 406 which includes records corresponding to related subjects. These relationships may be father-mother, sibling, twins, etc. Subjects may be part of a group that is being studied, e.g., a group with a congenital disease, or a toxic reaction to a particular drug. The groups are listed in a subject group table 408. Participation of subjects in groups is defined by a subject participation table 410 which lists all group memberships.

Samples and their attributes are listed in a sample table 412. Each sample has an associated sample type. The sample types are listed in a sample type table 414. Possible sample types include blood, urine, etc. Companies or institutions that provide samples are listed in a sample source table 416.

Database 102 provides an item table 418 that includes records for items. There are various types of items that correspond to different stages of the sample preparation process. An "item derivation" transforms an item of one type into an item of another type. The following table lists various item types and item derivation types for a representative embodiment.

|    | <u>Item Type</u> | <u>Derived from</u> | <u>by Item Derivation Type</u> |
|----|------------------|---------------------|--------------------------------|
| 30 | Sample           | other samples       | pooling                        |
|    | Sample           | other sample        | splitting                      |

|  |                 |   |
|--|-----------------|---|
| Extracted DNA                            | Sample          | DNA Extraction                                |
| Target (Sequences of interest amplified) | Extracted DNA   | PCR   |
| Fluorescently Labeled                    | Target          | Labeling                                      |
| 5 Target                                 |                 |   |
| Hybridized Chip                          | Labeled Target  | Hybridization (application of target-to chip) |
| Stained Hybridized Chip                  | Hybridized Chip | Staining                                      |

Item derivations are listed in an item derivation table 420. It should be noted that  
 10 derivations need not produce a change between item types. Each item derivation occurs  
 in accordance with a protocol that characterizes the step or steps in the derivation.  
 Protocols are listed in a protocol table 428. Each item derivation is performed by an  
 employee listed in employee table 432.

Unused chips are listed in a chip table 422. Hybridized chips (i.e., chips  
 15 that have had target applied) are listed in a hybridized chip table 424. A hybridized  
 sample map table 426 lists the relationships between hybridized chips and the samples  
 that have been applied to them.

Stained hybridized chips are scanned in a process referred to here as a  
 scan experiment. Scan experiments are listed in a scan experiment table 430. The scan  
 20 experiment occurs in accordance with a protocol listed in protocol table 428. The scan  
 experiment is performed by an employee listed in employee table 432.

Fig. 4B depicts further details of the data model for items and item  
 derivations. The various item types are listed in an item type table 434 and the various  
 item derivation types are listed in an item derivation type table 436. The relationships  
 25 between successive item types, e.g., sample and target are defined in an item type  
 derivation table 438. An item has associated attributes. For example, for a target,  
 database 102 may store the concentration, volume, location and/or remaining amount.  
 All item attributes are stored in an item attribute table 440. Item attributes may be  
 shared among multiple items. For example, a series of targets may all share a  
 30 preparation date. An item attribute item map table 442 implements a many-to-many

relationship between item attributes and items. The various types of item attributes such as preparer, preparation date, etc. are listed in an item attribute type table 444. Each item type has corresponding attribute types. Some attribute types are, however, shared among various item types. Accordingly, there is a many-to-many relationship among  
5 item attribute types and item types that is implemented by an item type map table 446.

The tables of Fig. 4B represent a powerfully general model of the sample preparation process. Changes in process steps that require changes in the type of information that should be stored may be implemented by changing and adding table contents rather than providing new tables or changing relationships among tables.

10 Fig. 4C depicts a detailed data model for storing information about protocols according to the present invention. Protocols as stored in protocol table 428 represent information about particular processes that have been performed including item derivations, analyses, and scan experiments. Each protocol has an associated protocol template. Protocol templates identify protocol types. For example, one protocol  
15 template may be a PCR template. All protocols associated with the PCR template identify parameters for performing a PCR procedure. Protocol templates are listed in a protocol template table 448. A parameter table 450 lists all the parameters and their values for all the protocols listed in protocol table 428. A parameter template table 452 lists the various parameter types along with default values. An examples of a parameter  
20 template would be a PCR reaction temperature. The parameter template would include a default value for this parameter. Parameter table 450 might then list many different PCR reaction temperature values that would be used by many different protocols. If a parameter value has not been modified by the user, it inherits the standard value of the associated parameter template. A parameter template set is a set of parameter templates  
25 that are used for a particular purpose, e.g., in association with protocols according to one or more protocol templates. Parameter template sets are listed in a parameter template set table 454. There are different types of parameter template set and these are listed in a parameter template set table 456. A mapping between parameter template sets and protocol templates is defined by a protocol template set map table 458.

30 Protocol templates may have associated lengthy verbal information about how to perform protocol steps. A protocol template document table 460 stores references to documents that include instructions for performing protocols.



As with the items, the data model for protocols defined by Fig. 4C is highly general and allows significant changes in the way item derivations, analyses, and experiments are performed without changing the underlying data model.

Referring again to Fig. 4A, there are tables to record information concerning the use of primers in PCR. A fragment table 462 lists all the sequence fragments investigated in conjunction with database 102. Associated with each fragment are one or more primer pairs used to amplify the fragment in a PCR process. A primer pair table 464 lists all the primer pairs including information about whether the primer pair actually worked to amplify the fragment. In order to develop the information about the effectiveness of primer pairs, there is a PCR table 466 that lists records identifying the outcome of multiple PCR operations. The individual PCR operations are identified by reference to item derivation table 420.

A single PCR operation may be used to amplify many different fragments and thus employ many different primer pairs. Of course, a single primer pair may be used in multiple PCR operations. There is therefore a many-to-many relationship between PCR operations and primer pairs that is recorded by a primer pair PCR map table 468. Information about individual primers is stored in a primer table 470. Also, each primer has an associated protocol in protocol table 428 that characterizes the primer preparation process. Information about primer orders is listed in a primer order table 472. Each primer order is to a vendor and the vendors are listed in a vendor table 474. Each primer order is made by an employee listed in employee table 432. A primer order design map table 476 implements a many-to-many relationship between primer orders and primers.

The data model described here thus preserves information about primers used in PCR reactions. One can improve results by using primers that have successfully amplified a given fragment in the past. Sometimes particular groups of primer pairs cannot be multiplexed together in the same PCR process. The information preserved here thus permits experimenters to make optimal use of expensive and time consuming PCR procedures.

It is also useful to preserve information about the chip production process and the origin of individual chips. A wafer table 478 lists wafers. When chips are produced, many chips are produced at the same time as part of a single wafer. Chip

table 422 stores references to wafer table 478 for each chip and the location of each chip on its wafer at production time. Sometimes there is analytic significance associated with the location of a chip on the wafer. Each wafer is produced as part of a lot and the identify of the lot for each wafer is recorded by wafer table 478 as a reference to a lot  
5 table 480 that lists each lot.

Fig. 4D depicts further details of tables pertaining to chip design that are preferably maintained within polymorphism database 102 according to one embodiment of the present invention. A tiling design table 482 lists tiling designs. Each tiling design represents the application of a particular tiling format to a sequence to be investigated.  
10 Tiling formats indicate probe orientation, probe length, and the position within a probe of a single nucleotide polymorphism being investigated. In a preferred embodiment, there may be very few tiling formats and they are listed in a tiling format table 484.

A particular tiling design includes many atom designs specifying the design of a single atom. In one embodiment, an atom is a group of typically four probes used to investigate a single base position with each probe hybridizing to a sequence including  
15 a different base at that position. Atom designs are listed in an atom design table 486. Records identifying the designs of individual probes are listed in a probe design table 488. A probe design role table 490 indicates the roles of probes listed in probe design table 488 in the atom designs of atom design table 486. For combinations of probe  
20 design and atom design, probe design role table 490 indicates which base the probe hybridizes to at the substitution position and whether the probe represents a match or a mismatch to the wild type.

A probe data table 492 gives the hybridization intensity values for particular probes designs as determined in particular scan experiments. Each record of  
25 the table also gives the number of pixels used to determine the intensity value and the standard deviation of intensity as measured among the pixels.

Figs. 4E-4G depict aspects of polymorphism database 102 related to analysis procedures and their results according to one embodiment of the present invention. An analysis table 494 lists analyses performed. An analysis generally refers  
30 to a non-trivial transformation of data. Records of analysis table 494 include references to protocol table 428 to specify parameters used for each analysis. Analyses may take as their input raw data or the results of previous analyses. An analysis dependency table

496 lists dependencies among analyses where one analysis depends on the data developed by another analysis. An analysis input table 498 lists inputs for analyses listed in analysis table 494.

On the right side of Fig. 4E are various tables used to support analyses.

5 A chip design sequence map table 500 correlates particular fragments with chip designs. A sequence position table 502 lists investigated sequence positions indicating their positions on a fragment. Records of sequence position table 502 reference a genomic sequence position table 504 which gives sequence positions in the genome rather than within individual fragments.

10 A scan experiment set table 506 lists sets of scan experiments. This allows for groupings of experiments for individuals or populations to serve as the basis for polymorphism analysis. A scan experiment used table 508 lists records indicating memberships of a scan experiment in a scan experiment set.

A tiling data table 510 lists records identifying tiling designs as  
15 implemented in particular chips measured by particular scan experiments. An atom data table 512 lists the intensities measured for particular sequence positions as measured in scan experiments identified by the tiling data records. A subject sequence position data table 514 lists combinations of sequence position and scan experiment.

A series of tables in Figs. 4E-4G correspond to different types of analysis  
20 that occur during the course of a polymorphism investigation. The types presented here are merely representative. A parallel series of tables provide the analysis results. A polymorphism analysis table 516 lists references to analysis table 494. The results of the performed polymorphism analyses are listed in a polymorphism position result table 518. A record of this table gives a result for a polymorphism analysis for a particular position  
25 as determined based on a particular set of scan experiments. In one embodiment the result is whether a particular mutation is certain, likely, possible, or not possible at the position. The result may also be that the reference is wrong.

A user polymorphism analysis table 520 lists user interpretations of results as listed in polymorphism position result table 518. The records of user polymorphism  
30 analysis table 520 are references to analysis table 494. The user interpretations themselves are stored in a user polymorphism analysis result table 522. Each result is a

likelihood of a particular mutation at a position as considered by a user plus an accompanying user comment.

5 A P-Hat analysis estimates the relative concentrations of wild type sequence and sequence having a particular mutation as determined in a particular scan experiment. A P-Hat analysis table 524 lists references to analysis table 494. An atom result table 526 gives estimates of the relative concentration along with upper and lower bounds and a maximum intensity. For heterozygous mutations, the estimates of relative concentration will cluster around 0.5. For homozygous mutations, the estimates should cluster around 1.0.

10 Base call analyses are determinations of the base at a particular position for a particular individual that may be based on more than one experiments. A base call analysis table 528 lists references to analysis table 494. A base call result table 530 lists the called bases for particular combinations of sequence position and subject.

15 A P-Hat grouping analysis determines a measure of likelihood that data in a set of scan experiments results from separate genotypes. P-hat grouping analyses are listed in a p-hat grouping analysis table 532 by reference to analysis table 494. P-hat grouping analysis results are listed in a mutation fraction result table 534. A group separation is given for various combinations of sequence position and scan experiment set.

20 A clustering analysis determines an alternative measure of likelihood that data in a set of scan experiments results from separate genotypes. Clustering analyses are listed in a clustering analysis table 536 by reference to analysis table 494. Clustering analysis results are listed in a clustering result table 538. A clustering factor is given for various combinations of sequence position and scan experiment set.

25 Fig. 4F shows tables which support normalization and footprint finding operations that support the analyses referred to in Fig. 4E. Hybridization intensity measurements made in scan experiments should be normalized over a set of scan experiments. The normalization should take into account differences in amplification level produced by different PCR processes.

30 Normalization is done by region of sequence. A normalization region analysis determines the boundaries of a region to be normalized. The determination of boundaries takes into account that different fragments of sequence are amplified by

different PCR procedures. A normalization region analysis table 540 lists normalization region analyses by reference to analysis table 494. A normalization region result table 542 lists the boundaries for each determined normalization region.

5        Normalization values for identified normalization regions are themselves determined by normalization analyses. Normalization analyses are listed in a normalization analysis table 544 by reference to analysis table 494. A normalization result table 546 lists the normalization values for regions.

10       A footprint analysis determines regions of sequence for which the hybridization intensity is elevated for the purposes of quality control. Footprint analyses are listed in a footprint analysis table 548 by reference to analysis table 494. Footprints are identified by sequence starting point and ending point in a particular scan experiment in a footprint table 550.

15       Fig. 4G depicts tables pertaining to measurement quality according to one embodiment of the present invention. A tiling data quality analysis determines the quality of results from a scan experiment. These analyses are listed in a tiling data quality analysis table 552 by reference to analysis table 494. Tiling data quality analysis results are listed in a tiling data quality result table 554. The results include an average hybridization intensity value for perfect match or mismatch probes. A wild type call rate gives the fraction of atom data where the probe corresponding to the reference base has the highest hybridization intensity. A wild type call rate of around 1.0 indicates good quality. Where the call rate is less than 0.75, the scan experiment should be rejected. An accept data field indicates whether the analysis indicates rejection or acceptance.

20       Where scan experiment measurements indicate two or more non-wild type bases within a probe length, this indicates a measurement problem for the affected region of sequence. These regions are identified by difficult region analyses listed in a difficult region analysis table 556 by reference to analysis table 494. A difficult region result table 558 lists the regions identified as being difficult.

25       Analysis dependency table 496 indicates interrelationships among the various analyses of Figs. 4E-4G. A footprint analysis may depend on a normalization analysis which may in turn depend on a normalization region analysis. A basecall analysis or PHatGrouping analysis may depend on an atom analysis. A polymorphism

analysis may depend on any of these analyses and/or a user polymorphism analysis and/or a clustering analysis.

Another aspect of the investigation of polymorphisms is seeking patent protection for identified polymorphisms. Fig. 4H shows tables of polymorphism database 102 related to efforts to seek patent protection according to one embodiment of the present invention. A polymorphism patent sequence table 560 lists sequences for which patent protection is sought. A patent application table 562 lists patent applications directed toward the protection of polymorphisms. A polymer patent application sequence map table 564 implements a many-to-many relationship between patent applications and sequences. A prior application table 566 lists relationships between patent applications and prior related patent applications. An attorney table 568 lists attorneys responsible for preparing patent applications listed in patent application table 562. A law firm table 570 lists the law firms to which the attorneys listed in attorney table 568 belong.

An employee group table 572 lists groups of inventors for the patent applications listed in table 562. Individual inventors are listed in employee table 432. An employee group map table 574 implements a many-to-many relationship between inventors and groups of inventors.

The data model of Fig. 4H greatly facilitates the process of securing patent protection for polymorphisms and thereby increases the commercial incentive for investigation of polymorphisms.

#### Database Contents

The contents of the tables introduced above will now be presented in greater detail in the following chart.

| TABLE      | FIELD | COMMENT |
|------------|-------|---------|
| tblSubject |       |         |

| TABLE | FIELD                    | COMMENT   |
|-------|--------------------------|---|
|       | SubjectId:INTEGER        | Identifies biological source of sample.                               |
|       | SpeciesID:INTEGER        | Species of subject.   |
|       | Name: VARCHAR2(20)       | Name of subject (anonimized for human subjects).                      |
|       | Gender: VARCHAR2(10)     | Gender of subject.  |
|       | Family: VARCHAR2(20)     | Family of subject (anonimized for human subjects).                    |
|       | Member:SMALLINT          | Position in family (father, mother, etc.).                            |
|       | Group_: VARCHAR2(20)     | Ethnic group.   |
|       | CellLineID: VARCHAR2(20) | Identifier for sample source not associated with particular organism. |

| TABLE               | FIELD                  | COMMENT   |
|---------------------|------------------------|---|
|                     | IsReference:SMALLINT   | Whether or not subject is in a group.                       |
| tblSpecies          |                        |   |
|                     | SpeciesId:INTEGER      | Species identifier.   |
|                     | Name:VARCHAR2(30)      | Name of species.  |
| SubjectRelationship |                        |   |
|                     | Subject1:INTEGER       | First subject in relationship.                              |
|                     | Subject2:INTEGER       | Second subject in relationship.                             |
|                     | Position:VARCHAR2(2)   | Nature of relationship.                                     |
| tblSubjectGroup     |                        |   |
|                     | GroupId:INTEGER        | Identifier of group of subjects (not same as ethnic group). |
|                     | GroupCode:VARCHAR2(20) | Code identifier for group.                                  |
|                     | Comments:LONG VARCHAR  | User comments on group.                                     |



| TABLE                   | FIELD                   | COMMENT                                 |
|-------------------------|-------------------------|---|
|                         | upsize_ts:DATE          | Creation date for group.                |
| tblSubjectParticipation |                         |   |
|                         | SubjectId:INTEGER       | Reference to subject table.             |
|                         | GroupId:INTEGER         | Reference to subject group table.       |
| tblSample               |                         |   |
|                         | SampleId:INTEGER        | Sample identifier.                      |
|                         | SubjectID:INTEGER       | Reference to subject table.             |
|                         | SampleSourceId:CHAR(18) | Institutional source of sample.         |
|                         | Code:VARCHAR2(20)       | Code representing individual subject.   |
|                         | Recipient:VARCHAR2(20)  | Person accepting sample.                |
|                         | Provider:VARCHAR2(20)   | Person or institution providing sample. |

| TABLE            | FIELD                     | COMMENT  |
|------------------|---------------------------|--|
|                  | DateReceived:DATE         | Date sample received.                                |
|                  | ProtocolId:INTEGER        | Reference to protocol table.                         |
|                  | SampleTypeId:INTEGER      | Reference to sample type table.                      |
| tblSampleType    |                           |  |
|                  | SampleTypeId:INTEGER      | Sample type identifier.                              |
|                  | Description:VARCHAR2(50)  | Description of sample type.                          |
| tblSample Source |                           |  |
|                  | SampleSourceId:CHAR(18)   | Identifier of institutional sample source.           |
|                  | ProviderName:VARCHAR2(20) | Name of individual or institutional sample provider. |
| Item             |                           |  |
|                  | ItemId:INTEGER            | Item identifier.                                     |
|                  | ItemTypeId:INTEGER        | Item type identifier.                                |
|                  | ItemName:VARCHAR2(50)     | Name of item.  |

| TABLE          | FIELD                         | COMMENT                              |
|----------------|-------------------------------|--------------------------------------|
| ItemDerivation |                               |                                      |
|                | Item1Id:INTEGER               | Derivation source.                   |
|                | Item2Id:INTEGER               | Derivation result.                   |
|                | EmployeeId:INTEGER            | Employee responsible for derivation. |
|                | DerivationTypeId:INTEGER      | Derivation type identifier.          |
|                | ProtocolId:VARCHAR2(18)       | Reference to protocol table.         |
|                | Date:DATE                     | Date of derivation.                  |
| tblChip        |                               |                                      |
|                | ChipId:INTEGER                | Rename reference to item table.      |
|                | ChipDesignPlacementId:INTEGER | Placement on wafer.                  |
|                | LocationId:INTEGER            | Location of chip.                    |
|                | WaferId:INTEGER               | Wafer the chip was on.               |
| tblHybedChip   |                               |                                      |

| TABLE             | FIELD                      | COMMENT  |
|-------------------|----------------------------|--|
|                   | HybedChipId:INTEGER        | Rename reference to item table.                            |
|                   | SubjectID:INTEGER          | Reference to subject table.                                |
|                   | ProtocolId:INTEGER         | Reference to protocol table.                               |
|                   | Repetition:SMALLINT        | Refers to number of times chip has been washed and reused. |
| tblHybSampleMap   |                            |  |
|                   | ItemId:INTEGER             | Reference to item table.                                   |
| Protocol          |                            |  |
|                   | ProtocolId:INTEGER         | Protocol identifier.                                       |
|                   | ProtocolTemplateId:INTEGER | Protocol template identifier.                              |
|                   | Name:VARCHAR2(100)         | Name of protocol.  |
| tblScanExperiment |                            |  |

| TABLE | FIELD                       | COMMENT   |
|-------|-----------------------------|---|
|       | ScanExptId:INTEGER          | Scan experiment identifier.                                 |
|       | ItemId:INTEGER              | Reference to item table.                                    |
|       | ScanCode:VARCHAR2(25)       | File for scan results.                                      |
|       | ProtocolId:INTEGERP         | Reference to protocol table.                                |
|       | ScanRatingId:INTEGER        | Assessment of scan quality.                                 |
|       | ExperimenterId:INTEGER      | Experimenter identifier.                                    |
|       | Date:DATE                   | Date of experiment.   |
|       | ConversionTool:VARCHAR2(50) | Program used to convert from scan image to intensities.     |
|       | ConversionDate:DATE         | Date of conversion.   |
|       | ScanStatus:VARCHAR2(50)     | whether or not scan image has been converted to intensities |
|       | Comments:LONG VARCHAR       | Comments.   |

| TABLE              | FIELD                      | COMMENT   |
|--------------------|----------------------------|---|
|                    |                            |   |
| Employee           |                            |   |
|                    | EmployeeId:INTEGER         | Employee identifier.                            |
|                    | EmployeeCode:VARCHAR2(5)   | Code for employee                               |
|                    | FName: VARCHAR2(20)        | First name of employee.                         |
|                    | MName: VARCHAR2(20)        | Middle name of employee.                        |
|                    | LName: VARCHAR2(20)        | Last name of employee.                          |
| ItemType           |                            |   |
|                    | ItemId:INTEGER             | Item type identifier.                           |
|                    | ItemTypeName: VARCHAR2(30) | Name of item type.                              |
|                    | FormName: VARCHAR2(100)    | Reference to user interface form for item type. |
| ItemDerivationType |                            |   |
|                    | DerivationTypeId:INTEGER   | Derivation type identifier.                     |

| TABLE                | FIELD                        | COMMENT                                 |
|----------------------|------------------------------|---|
|                      | DerivationType: VARCHAR2(50) | Description of derivation type.         |
| ItemTypeDerivation   |                              |   |
|                      | NextItemId: INTEGER          | Result type of derivation.              |
|                      | ItemTypeId: INTEGER          | Source type of derivation.              |
| ItemAttribute        |                              |   |
|                      | itemAttributeId: INTEGER     | Item attribute identifier.              |
|                      | ItemAttributeTypeId: INTEGER | Reference to item attribute type table. |
|                      | Attribute: VARCHAR2(50)      | Attribute value.                        |
| ItemAttributeItemMap |                              |   |
|                      | ItemAttributeId: INGEGER     | Reference to item attribute table.      |
|                      | ItemId: INTEGER              | Reference to item table.                |
| ItemAttributeType    |                              |   |
|                      | ItemAttributetypeId: INTEGER | Item attribute identifier.              |

| TABLE            | FIELD                           | COMMENT   |
|------------------|---------------------------------|---|
|                  | ItemAttributeName: VARCHAR2(30) | Name of item attribute type.                            |
| ItemTypeMap      |                                 |   |
|                  | ItemAttributeTypeId: INTEGER    | Reference to item attribute type table.                 |
|                  | ItemTypeId: INTEGER             | Reference to item type table.                           |
| ProtocolTemplate |                                 |   |
|                  | ProtocolTemplateId: INTEGER     | Protocol template identifier.                           |
|                  | Name: VARCHAR2(100)             | Name of protocol template.                              |
|                  | DateCreated: DATE               | Date protocol template created.                         |
|                  | FormName: VARCHAR2(50)          | Name of the electronic form used for protocol template. |
| Parameter        |                                 |   |
|                  | ParameterId: INTEGER            | Parameter identifier.                                   |



| TABLE             | FIELD                       | COMMENT                                    |
|-------------------|-----------------------------|--|
|                   | ParameterTemplateId:INTEGER | Reference to parameter template table.     |
|                   | Value:VARCHAR2(20)          | Value of parameter.                        |
|                   | ProtocolID:INTEGER          | Reference to protocol table.               |
| ParameterTemplate |                             |  |
|                   | ParameterTemplateId:INTEGER | Parameter template identifier.             |
|                   | Name:VARCHAR2(100)          | Name of parameter template.                |
|                   | ParamTemplateSetId:INTEGER  | Reference to parameter template set table. |
|                   | StandardValue:VARCHAR2(100) | Default value for parameter.               |
| ParamTemplateSet  |                             |  |
|                   | ParamTemplateSetId:INTEGER  | Parameter template set identifier.         |

| TABLE                   | FIELD                      | COMMENT   |
|-------------------------|----------------------------|---|
|                         | TypeId:INTEGER             | Renamed reference to parameter template set type table. |
|                         | Name:VARCHAR2(20)          | Name of parameter template set.                         |
| ParamTemplateSetType    |                            |   |
|                         | ParamTempSetTypeId:INTEGER | Parameter template set type identifier.                 |
|                         | Description: VARCHAR2(50)  | Description of parameter template set type.             |
| ParameterTemplateSetMap |                            |   |
|                         | ProtocolTemplateId:INTEGER | Reference to protocol template table.                   |
|                         | ParamTemplateSetId:INTEGER | Reference to parameter template set table.              |
| ProtocolTemplateDoc     |                            |   |

| TABLE      | FIELD                        | COMMENT                                      |
|------------|------------------------------|--|
|            | ProtocolDocId:INTEGER        | Protocol Template document identifier.       |
|            | ProtocolTemplateId:INTEGER   | Reference to protocol template table.        |
|            | Name:VARCHAR2(100)           | Name of protocol template.                   |
|            | PathAndFileName:VARCHAR2(50) | File name for protocol template document.    |
|            | AuthorName:INTEGER           | Author of protocol template document.        |
|            | CreationDate:DATE            | Creation Date of protocol template document. |
| tbFragment |                              |  |
|            | FragmentId:INTEGER           | Fragment identifier.                         |
|            | ChipSequence:LONG VARCHAR    | Sequence of fragment.                        |

| TABLE         | FIELD                  | COMMENT  |
|---------------|------------------------|--|
|               |                        |  |
|               | Code: VARCHAR2(50)     | Code representing fragment.                          |
| tblPrimerPair |                        |  |
|               | PrimerPairId: INTEGER  | Identifier for primer pair.                          |
|               | LeftPrimerId: INTEGER  | Left primer identifier.                              |
|               | RightPrimerId: INTEGER | Right primer identifier.                             |
|               | PCRSize: INTEGER       | length of amplified fragment                         |
|               | Worked: SMALLINT       | Whether or not pair successfully amplified fragment. |
|               | FragmentId: INTEGER    | Reference to fragment table.                         |
| tblPCR        |                        |  |
|               | Item1Id: INTEGER       | First part of reference to item derivation table.    |

| TABLE           | FIELD                   | COMMENT  |
|-----------------|-------------------------|--|
|                 | Item2Id:INTEGER         | Second part of reference to item derivation table. |
|                 | Reactionworked:SMALLINT | Whether or not PCR reaction worked.                |
| PrimePairPCRMap |                         |  |
|                 | PrimerPairId:INTEGER    | Reference to primer pair table.                    |
|                 | Item1ID:INTEGER         | First part of referenced item derivation table.    |
|                 | Item2Id:INTEGER         | Second part of referenced item derivation table.   |
| tblPrimer       |                         |  |
|                 | PrimerId:INTEGER        | Primer identifier.                                 |
|                 | ProtocolId:INTEGER      | Reference to protocol table.                       |
|                 | OligoSeq:VARCHAR2(35)   | Sequence of primer.                                |

| TABLE          | FIELD                  | COMMENT                         |
|----------------|------------------------|---------------------------------|
|                | Position:INTEGER       | Position of primer on fragment. |
|                | Length:INTEGER         | Length of primer.               |
|                | MeltingTemp:INTEGER    | Melting temperature of primer.  |
|                | Direction:VARCHAR2(20) | Direction (forward or reverse). |
| tblPrimerOrder |                        |                                 |
|                | OrderId:INTEGER        | Order identifier.               |
|                | EmployeeId:INTEGER     | Employee who made order.        |
|                | VendorId:INTEGER       | Vendor for order.               |
|                | OrderDate:DATE         | Date of order.                  |
|                | Owner: VARCHAR2(50)    | Name of employee making order.  |
|                | Vendor: VARCHAR2(50)   | Name of vendor.                 |
| tblVendor      |                        |                                 |

| TABLE                   | FIELD                     | COMMENT                     |
|-------------------------|---------------------------|-----------------------------|
|                         | VendorId:INTEGER          | Vendor identifier.          |
|                         | Vendor: VARCHAR2(50)      | Name of vendor.             |
|                         | PhoneNumber: VARCHAR2(15) | Phone number of vendor.     |
|                         | FaxNumber: VARCHAR2(15)   | Fax Number of vendor.       |
|                         | Address: VARCHAR2(50)     | Address of vendor.          |
|                         | City: VARCHAR2(50)        | City of vendor.             |
|                         | State: VARCHAR2(50)       | State of vendor.            |
|                         | Zip: VARCHAR2(50)         | Zip code of vendor.         |
| tblPrimerOrderDesignMap |                           |                             |
|                         | PrimerId:INTEGER          | Reference to primer table.  |
|                         | OrderId:INTEGER           | Reference to order table.   |
| tblWafer                |                           |                             |
|                         | WaferId:INTEGER           | Wafer identifier.           |
|                         | LotId:INTEGER             | Lot to which wafer belongs. |

| TABLE            | FIELD                     | COMMENT                               |
|------------------|---------------------------|---------------------------------------|
|                  | Code:VARCHAR2(8)          | Code for wafer.                       |
|                  | SynthesisDate_delete:DATE | Synthesis date for wafer.             |
|                  | Released:DATE             | Date wafer available.                 |
|                  | Done:SMALLINT             | Whether wafer production is complete. |
|                  |                           |                                       |
|                  | ExpirationDate:DATE       | Expiration date of wafer.             |
|                  | ExpectedLife:CHAR(18)     | Expected useful life of wafer.        |
| tblLot           |                           |                                       |
|                  | LotId:INTEGER             | Lot identifier.                       |
|                  | WaferDesignId:INTEGER     | Identifier for wafer design.          |
|                  | LotNumber:VARCHAR2(12)    | Lot number.                           |
|                  | WaferPN:VARCHAR2(50)      | Part number for wafer.                |
| tblTiling Design |                           |                                       |
|                  | TilingDesignID:INTEGER    | Tiling design identifier.             |



| TABLE            | FIELD                          | COMMENT   |
|------------------|--------------------------------|---|
|                  | ChipDesignSequenceMapID:NUMBER | Reference to chip design sequence map.  |
|                  | TilingFormatID:INTEGER         | Reference to tiling format table.   |
|                  | UnitNumber:INTEGER             | 1 for sense, 0 for antisense  |
|                  | AtomOffset:INTEGER             | # to add to translate atom position in tiling to atom position in chip design |
| tblTiling Format |                                |   |
|                  | TilingFormatID:INTEGER         | Tiling format identifier  |
|                  | Orientation:CHAR(18)           | Orientation for tiling.   |
|                  | ProbeLength:SMALLINT           | Length of probes.   |
|                  | SubstitutionPosition:SMALLINT  | Substitution position for mutation base in probes.                            |
| tblAtomDesign    |                                |   |

| TABLE              | FIELD                  | COMMENT                           |
|--------------------|------------------------|-----------------------------------|
|                    | AtomDesignId:NUMBER    | Atom design identifier.           |
|                    | TilingDesignID:INTEGER | Reference to tiling design table. |
|                    | Position:INTEGER       | Position of atom in sequence.     |
| tblProbeDesign     |                        |                                   |
|                    | ProbeDesignID:NUMBER   | Probe design identifier.          |
|                    | ChipDesignId:INTEGER   | Reference to chip design.         |
|                    | x:SMALLINT             | x position of probe.              |
|                    | y:SMALLINT             | y position of probe.              |
| tblProbeDesignRole |                        |                                   |
|                    | ProbeDesignID:NUMBER   | Reference to probe design table.  |
|                    | AtomDesignID:NUMBER    | Reference to atom design table.   |

| TABLE        | FIELD                 | COMMENT  |
|--------------|-----------------------|--|
|              | Substitution:CHAR(18) | Substitution position in probe design.           |
|              | Mismatches:NUMBER     | Whether probe is match or mismatch.              |
| tblProbeData |                       |  |
|              | ProbeDesignID:NUMBER  | Reference to probe design table.                 |
|              | ScanExptID:INTEGER    | Reference to scan experiment table.              |
|              | Intensity:FLOAT       | Measured hybridization intensity for probe.      |
|              | NPixels;NUMBER        | Number of pixels used for intensity calculation. |
|              | StDev:NUMBER          | Standard deviation for pixels.                   |
| tblAnalysis  |                       |  |

| TABLE                 | FIELD                     | COMMENT                                   |
|-----------------------|---------------------------|---|
|                       | AnalysisId:INTEGER        | Analysis identifier.                      |
|                       | AnalysisVersionID:INTEGER | Reference to version of analysis.         |
|                       | ProtocolID:INTEGER        | Reference to protocol table.              |
|                       | DatePerformed:DATE        | Date analysis performed.                  |
|                       | NeedsUpdate:NUMBER        | Whether analysis is current.              |
| tblAnalysisDependency |                           |   |
|                       | ParentAnalysisId:INTEGER  | Analysis providing input.                 |
|                       | SubAnalysisId:INTEGER     | Analysis receiving input.                 |
|                       | Role:VARCHAR2(20)         | Role of data provided by parent analysis. |
| TblAnalysisInput      |                           |   |
|                       | AnalysisinputID:INTEGER   | Analysis input identifier.                |

| TABLE                    | FIELD                          | COMMENT   |
|--------------------------|--------------------------------|---|
|                          | AnalysisId:INTEGER             | Analysis receiving input.   |
|                          | Inputtype:VARCHAR2(20)         | Type of input.  |
|                          | ObjectID:INTEGER               | Reference to input data.  |
| tblChipDesignSequenceMap |                                |   |
|                          | ChipDesignSequenceMapID:NUMBER | Chip design sequence map identifier.  |
|                          | FragmentID:INTEGER             | Reference to fragment table.  |
|                          | ChipDesignId:INTEGER           | Chip design identifier.   |
|                          | AtomOffset:NUMBER              | # to add to translate atom position in tiling to atom position in chip design |
| tblSequencePosition      |                                |   |
|                          | SequencePositionID:NUMBER      | Sequence position identifier.   |
|                          | ChipDesignSequenceMapID:NUMBER | Reference to chip design sequence map table.                                  |

| TABLE                      | FIELD                             | COMMENT                                       |
|----------------------------|-----------------------------------|---|
|                            | Position:NUMBER                   | Position in fragment.                         |
|                            | GenomicSequencePositionID:INTEGER | Reference to genomic sequence position table. |
|                            | RefBase:INTEGER                   | Reference base.                               |
| tblGenomicSequencePosition |                                   |   |
|                            | GenomicSequencePositionID:INTEGER | Genomic sequence position identifier.         |
| tblScanExperimentSet       |                                   |   |
|                            | ScanExperimentSetID:NUMBER        | Scan experiment set identifier.               |
| tbsScanExperimentUsed      |                                   |   |
|                            | ScanExptID:INTEGER                | Reference to scan experiment table.           |
|                            | ScanExperimentSetID:NUMBER        | Reference to scan experiment set table.       |
| tblTilingData              |                                   |   |
|                            | TilingDataID:NUMBER               | Tiling data identifier.                       |
|                            | ScanExptID:INTEGER                | Reference to scan experiment table.           |
|                            | TilingDesignID:INTEGER            | Reference to tiling design table.             |
| tblAtomData                |                                   |   |

5

| TABLE                      | FIELD                             | COMMENT                                       |
|----------------------------|-----------------------------------|---|
|                            | AtomDataID:INTEGER                | Atom data identifier.                         |
|                            | TilingDataID:NUMBER               | Reference to tiling data table.               |
|                            | SubjectSequencePositionID:INTEGER | Reference to subject sequence position table. |
| tblSubjectSequencePosition |                                   |   |
|                            | SubjectSequencePositionID:INTEGER | Subject sequence position identifier.         |
|                            | SubjectID:INTEGER                 | Reference to subject table.                   |
|                            | SequencePositionID:NUMBER         | Reference to sequence position table.         |
| tblPolymorphismAnalysis    |                                   |   |
|                            | AnalysisId:INTEGER                | Reference to analysis table.                  |
| tblPolyPositionResult      |                                   |   |
|                            | AnalysisId:INTEGER                | Reference to analysis table.                  |
|                            | PolyPositionID:INTEGER            | Polymorphism position identifier.             |
|                            | ScanExperimentSetID:NUMBER        | Reference to scan experiment set table.       |

| TABLE                     | FIELD                      | COMMENT  |
|---------------------------|----------------------------|--|
|                           | PolyPositionTypeID:INTEGER | Refers to possibility of polymorphism at position, e.g., certain, likely, possible, mismatch (reference is wrong). |
|                           | WTBase:CHAR(18)            | Wild type base at position.  |
|                           | MuBase:INTEGER             | Mutation base at position.   |
| tblUserPolyanalysis       |                            |  |
|                           | AnalysisId:INTEGER         | Reference to analysis table.   |
| tblUserPolyanalysisResult |                            |  |
|                           | AnalysisId:INTEGER         | Reference to analysis table.   |
|                           | SequencePositionID:NUMBER  | Reference to sequence position table.  |
|                           | ScanExperimentSetID:NUMBER | Reference to scan experiment set table.  |
|                           | PolyPositionTypeID:INTEGER | See polymorphism position result table.  |



| TABLE           | FIELD                      | COMMENT   |
|-----------------|----------------------------|---|
|                 | UserComment: VARCHAR2(256) | User comment done polymorphism analysis.        |
| tblAtomanalysis |                            |   |
|                 | AnalysisId: INTEGER        | Reference to analysis table.                    |
|                 |                            |   |
| tblAtomResult   |                            |   |
|                 | AnalysisId: INTEGER        | Reference to analysis table.                    |
|                 | AtomDataID: INTEGER        | Reference to atom data table.                   |
|                 | PHat: FLOAT                | Relative concentration of mutant and wild type. |
|                 | PHatUpperbound: FLOAT      | Upperbound for relative concentration.          |
|                 | PHatLowerbound: FLOAT      | Lowerbound for relative concentration.          |
|                 | MaxIntensity: FLOAT        | Maximum measured intensity for atom.            |
|                 | WTIntensity: FLOAT         | Measured wild type intensity.                   |

| TABLE               | FIELD                             | COMMENT  |
|---------------------|-----------------------------------|--|
|                     | MutIntensity:FLOAT                | Measured mutation intensity.   |
|                     | LocalWTCallRate:FLOAT             | rate at which atoms associated with surrounding sequence call reference base |
|                     | IntensityRatio:FLOAT              | Ratio of intensity of wild type probe over intensity of mutation probe.      |
| tblBaseCallAnalysis |                                   |  |
|                     | AnalysisId:INTEGER                | Reference to analysis table.   |
| tblBaseCallResult   |                                   |  |
|                     | AnalysisId:INTEGER                | Reference to analysis table.   |
|                     | SubjectSequencePositionID:INTEGER | Reference to sequence position table.  |
|                     | ScanExperimentSetID:NUMBER        | Reference to skin experiments set table.                                     |
|                     | CalledBase:VARCHAR2(1)            | Base called for subject based on experiment set.                             |

| TABLE                          | FIELD                         | COMMENT  |
|--------------------------------|-------------------------------|--|
|                                | SuggestCheck:NUMBER           | Used to indicate whether this sample should be used for resequencing |
| tblClusteringAnalysis          |                               |  |
|                                | AnalysisId:INTEGER            | Reference to analysis table.   |
| tblClusteringResult            |                               |  |
|                                | AnalysisId:INTEGER            | Reference to analysis table.   |
|                                | SequencePositionID:NUMBER     | Reference to sequence position table.                                |
|                                | ScanExperimentSetID:NUMBER    | Reference to scan experiment set table.                              |
|                                | ClusteringFactor:FLOAT        | Result of clustering analysis.                                       |
| tblNormalizationRegionAnalysis |                               |  |
|                                | AnalysisId:INTEGER            | Reference to analysis table.   |
| tblNormalizationRegion         |                               |  |
|                                | NormalizationRegionID:INTEGER | Normalization region identifier.                                     |
|                                | AnalysisId:INTEGER            | Reference to analysis table.   |

| TABLE                    | FIELD                               | COMMENT  |
|--------------------------|-------------------------------------|--|
|                          | ChipDesignSequenceMapID:NUMBER      | Reference to chip design sequence map table.         |
|                          | NumberScanExpt.Set                  | Reference to scan experiment set table.              |
|                          | RegionEnd:INTEGER                   | Indication of end of the normalization region.       |
|                          | RegionStart:INTEGER                 | Indication of beginning of the normalization region. |
| tblNormalizationAnalysis |                                     |  |
|                          | AnalysisId:INTEGER                  | Reference to analysis table.                         |
| tblNormalizationResult   |                                     |  |
|                          | NormalizationResultID:INTEGER       | Normalization result identifier.                     |
|                          | AnalysisId:INTEGER                  | Reference to analysis table.                         |
|                          | TilingDataID:INTEGER                | Reference to tiling data table.                      |
|                          | NormalizationRegionResultID:INTEGER | Reference to normalization result.                   |
|                          | NormalizationValue:NUMBER           | Value used for normalization.                        |

| TABLE                        | FIELD                          | COMMENT  |
|------------------------------|--------------------------------|--|
|                              | DataOK:NUMBER                  | Indication whether normalization result is usable. |
| tblFootprintAnalysis         |                                |  |
|                              | AnalysisId:INTEGER             | Reference to analysis table.                       |
| tblFootprint                 |                                |  |
|                              | FootprintID:NUMBER             | Footprint identifier.                              |
|                              | AnalysisId:INTEGER             | Analysis identifier.                               |
|                              | ChipDesignSequenceMapID:NUMBER | Reference to chip design sequence map table.       |
|                              | ScanExperimentSetID:NUMBER     | Reference to scan experiment set table.            |
|                              | FFStart:NUMBER                 | Start of footprint and sequence.                   |
|                              | FPend:NUMBER                   | End of footprint and sequence.                     |
| tblTilingDataQualityAnalysis |                                |  |
|                              | AnalysisId:INTEGER             | Reference to analysis table.                       |
| tbltilingDataQualityResult   |                                |  |
|                              | TilingDataID:NUMBER            | Reference to tiling data table.                    |

| TABLE                       | FIELD                          | COMMENT  |
|-----------------------------|--------------------------------|--|
|                             | AnalysisId:INTEGER             | Reference to analysis table.   |
|                             | AvgWTIntensity:NUMBER          | Average wild type intensity.   |
|                             | WTCallRate:NUMBER              | Fraction of atoms where brightest of probes is one with reference space. |
|                             | AcceptData:INTEGER             | Whether data is of acceptable quality.                                   |
| tblDifficult Regionanalysis |                                |  |
|                             | AnalysisId:INTEGER             | Reference to analysis table.   |
| tblDifficultRegionResult    |                                |  |
|                             | ScanExptId:INTEGER             | Reference to scan experiment table.                                      |
|                             | AnalysisId:INTEGER             | Reference to analysis table.   |
|                             | ChipDesignSequenceMapID:NUMBER | Reference to chip design sequence map table.                             |
|                             | RgnStart:NUMBER                | Beginning of difficult region in sequence.                               |
|                             | RgnEnd: NUMBER                 | End of difficult region in sequence.                                     |

| TABLE            | FIELD                     | COMMENT   |
|------------------|---------------------------|---|
|                  | Reason:INTEGER            | Code indicating reason for difficult region, e.g., two or more non-wild type bases and less than a probe length.q |
| tblPolyPatentSeq |                           |   |
|                  | PolyPatentSeqId:NUMBER    | Polymorphism sequence identifier.   |
|                  | Polyscreen:VARCHAR2(50)   | reference to internal grouping of polymorphisms   |
|                  | FragmentCode:VARCHAR2(50) | Fragment sequence found in  |
|                  | Position:LONG             | Position of polymorphism.   |
|                  | RefAllele:CHAR(2)         | Wild type base at position.   |
|                  | FreqP:FLOAT               | Frequency of wild type.   |
|                  | AltAllele:CHAR(2)         | Mutation base at position.  |
|                  | FreqQ:FLOAT               | Frequency of mutation base.   |
|                  | Heterozygosity:FLOAT      | Heterozygosity value.   |

| TABLE        | FIELD                       | COMMENT   |
|--------------|-----------------------------|---|
|              | SequenceTag: VARCHAR2(50)   | Sequence containing polymorphism including ambiguity code at polymorphism position. |
|              | GeneName: VARCHAR2(50)      | Name of gene.   |
|              | ChromosomeNum: VARCHAR2(20) | Chromosome number.  |
|              | ChromosomeLoc: VARCHAR2(20) | Location of gene on chromosome.   |
|              | ForwardPrimer: VARCHAR2(50) | Identifier for forward primer used to implement fragment.                           |
|              | ReversePrimer: VARCHAR2(50) | Identifier of primer used to amplify fragment.                                      |
| tblPatentApp |                             |   |
|              | PatentAppId: NUMBER         | Patent application identifier.  |
|              | GroupId: NUMBER             | Reference to employee group table.  |
|              | AttorneyId: NUMBER          | Reference to attorney table.  |
|              | DocketNum: VARCHAR2(30)     | Docket number for patent application.   |



| TABLE               | FIELD                        | COMMENT  |
|---------------------|------------------------------|--|
|                     | FilingDate:DATE              | Filing date for filing application.  |
|                     | Classification:VARCHAR2(30)  | Patent office classification for patent application.                       |
|                     | SerialNumber:VARCHAR2(50)    | Serial number assigned by patent office.                                   |
|                     | CountryCode:VARCHAR2(50)     | Country in which patent application was filed.                             |
|                     | InventionTitle:VARCHAR2(100) | Title for patent application   |
| tblPolyPatentSeqMap |                              |  |
|                     | PatentAppId:NUMBER           | Reference to patent application table.                                     |
|                     | PolyPatentSeqId:NUMBER       | Reference to polymorphism patent sequence table.                           |
| tblPriorApp         |                              |  |
|                     | PriorAppId:NUMBER            | Reference to related prior patent application in patent application table. |

| TABLE       | FIELD                        | COMMENT   |
|-------------|------------------------------|---|
|             | AppId:NUMBER                 | Reference to application to which prior application is related. |
| tblAttorney |                              |   |
|             | AttorneyId:NUMBER            | Attorney identifier.  |
|             | LawFirmId:NUMBER             | Law firm where attorney works.                                  |
|             | FirstName:VARCHAR2(20)       | First name of attorney.   |
|             | MiddleName:VARCHAR2(5)       | Middle name of attorney.  |
|             | LastName:VARCHAR2(30)        | Last name of attorney.  |
|             | RegistrationNum:VARCHAR2(25) | Patent office registration number of attorney.                  |
| tblLawFirm  |                              |   |
|             | LawFirmId:NUMBER             | Law firm identifier.  |
|             | Company:VARCHAR2(100)        | Name of law firm.   |
|             | Address:VARCHAR2(100)        | Address of law firm.  |
|             | City:VARCHAR2(30)            | City address of law firm.                                       |

| TABLE             | FIELD  | COMMENT   |
|-------------------|--|---|
|                   | State: VARCHAR2(20)                          | State address of law firm.                          |
|                   | ZipCode: VARCHAR2(15)                        | Zip Code of law firm.                               |
|                   | Country: VARCHAR2(15)                        | Country of law firm.                                |
|                   | Telephone: VARCHAR2(30)<br>Fax: VARCHAR2(30) | Telephone number of law firm.                       |
|                   | TELEX: VARCHAR2(20)                          | Facsimile number of law firm.                       |
|                   |  | Telex number of law firm.                           |
| tblEmployeeGroup  |  |   |
|                   | GroupId: NUMBER                              | Identifier for inventor group.                      |
|                   | GroupName: VARCHAR2(50)                      | Name of inventor group.                             |
|                   | Comments: VARCHAR2(50)                       | Comments.   |
|                   | GroupList: VARCHAR2(255)                     | Written out list of inventor names.                 |
| tblEmployeeGrpMap |  |   |
|                   | EmployeeId: INTEGER                          | Reference to employee table for inventor/employees. |
|                   | GroupId: NUMBER                              | Reference to inventor group table.                  |

It is understood that the examples and embodiments described herein are for illustrative purposes only and that various modifications or changes in light thereof will be suggested to persons skilled in the art and are to be included within the spirit and purview of this application and scope of the appended claims. For example, tables may be deleted, contents of multiple tables may be consolidated, or contents of one or more tables may be distributed among more tables than described herein to improve query speeds and/or to aid system maintenance. Also, the database architecture and data models described herein are not limited to biological applications but may be used in any application. All publications, patents, and patent applications cited herein are hereby incorporated by reference.

**WHAT IS CLAIMED IS:**

1                   1.     A computer-readable storage medium having stored  
2     thereon:  
3                   an item table listing a plurality of item records identifying items;  
4     an item attribute table listing a plurality of item attribute records identifying  
5     attributes of said items; and  
6     wherein there is a many-to-many relationship between item records and item  
7     attribute records.

1                   2.     The computer-readable storage medium of claim 1  
2     wherein  
3                   an item attribute item map table implements said many-to-many  
4     relationship between item records and item attribute records, said item attribute  
5     item map table listing a plurality of map records identifying both a particular  
6     item attribute and a particular item.

1                   3.     The computer-readable storage medium of claim 1 having  
2     further stored thereon:  
3                   an item derivation table listing a plurality of item derivation  
4     records identifying transformations between ones of said items used in  
5     biological analysis.

1                   4.     The computer-readable storage medium of claim 3 having  
2     further stored thereon:  
3                   a protocol table listing a plurality of protocol records specifying  
4     parameters of said transformation.

1                   5.     The computer-readable storage medium wherein said items  
2     are used in a biological analysis.

1                   6.     The computer-readable storage medium of claim 1  
2     wherein said biological analysis comprises a polymorphism analysis.

1                   7.     A computer-readable storage medium having stored  
2     thereon:  
3                   an atom result table listing a plurality of atom result records,  
4     specifying relative wild-type and mutant sequence concentrations in targets; and  
5     a subject sequence position table listing a plurality of subject sequence position  
6     records, specifying combinations of subjects from whom said targets are derived  
7     and sequence positions, each said atom result record being associated with one  
8     or more atom result records.

1                   8.     The computer-readable storage medium of claim 7  
2     wherein said atom result records further specify upper and lower bounds for  
3     said concentrations.

1                   9.     The computer-readable storage medium of claim 7 having  
2     further stored thereon:  
3                   a subject table listing subject records specifying said subjects.

1                   10.    A computer-readable storage medium having stored  
2     thereon:  
3                   a polymorphism table listing polymorphism sequence records  
4     specifying sequences known to contain polymorphisms; and  
5                   a patent application table listing patent application records  
6     specifying one or more polymorphisms specified by said polymorphism  
7     sequence records.

1                   11.    The computer-readable storage medium of claim 10  
2     wherein said polymorphism sequence records specify for each one of said  
3     polymorphisms a polymorphism position, a reference allele, and a base allele.

1                   12.    The computer-readable storage medium of claim 11  
2    wherein said polymorphism sequence records further specify for each one of  
3    said polymorphisms a measured heterozygosity.

1                   13.    A computer-implemented method comprising:  
2    creating an item table listing a plurality of item records identifying items used in  
3    biological analysis; and  
4                    creating an item attribute table listing a plurality of item attribute  
5    records identifying attributes of said items; and  
6    wherein there is a many-to-many relationship between item records and item  
7    attribute records.

1                   14.    The computer-implemented method of claim 13 further  
2    comprising the step of:  
3                    creating an item attribute item map table implements said many-  
4    to-many relationship between item records and item attribute records, said item  
5    attribute item map table listing a plurality of map records identifying both a  
6    particular item attribute and a particular item.

1                   15.    The computer-implemented method of claim 13  
2    comprising:  
3                    an item derivation table listing a plurality of item derivation  
4    records identifying transformations between ones of said items used in  
5    biological analysis.

1                   16.    The computer-implemented method of claim 15 further  
2    comprising:  
3                    creating a protocol table listing a plurality of protocol records  
4    specifying parameters of said transformation.

1                   17.    The computer-implemented method of claim 13 wherein  
2    said biological analysis comprises a polymorphism analysis.

1                   18.    A computer-implemented method comprising:  
2    creating an atom result table listing a plurality of atom result records, specifying  
3    relative wild-type and mutant sequence concentrations in targets; and  
4    creating a subject sequence position table listing a plurality of subject sequence  
5    position records, specifying combinations of subjects from whom said targets  
6    are derived and sequence positions, each said atom result record being  
7    associated with one or more atom result records.

1                   19.    The computer-implemented method of claim 18 wherein  
2    said atom result records further specify upper and lower bounds for said  
3    concentrations.

1                   20.    The computer-implemented method of claim 18 further  
2    comprising:  
3                   creating a subject table listing subject records specifying said  
4    subjects.

1                   21.    A computer-implemented method comprising:  
2                   creating a polymorphism table listing polymorphism sequence  
3    records specifying sequences known to contain polymorphisms; and  
4                   creating a patent application table listing patent application records  
5    specifying one or more polymorphisms specified by said polymorphism  
6    sequence records.

1                   22.    The computer-implemented method of claim 21 wherein  
2    said polymorphism sequence records specify for each one of said  
3    polymorphisms a polymorphism position, a reference allele, and a base allele.



- 1                   23.    The computer-implemented method of claim 22 wherein  
2    said polymorphism sequence records further specify for at least one of said  
3    polymorphisms a measured heterozygosity.

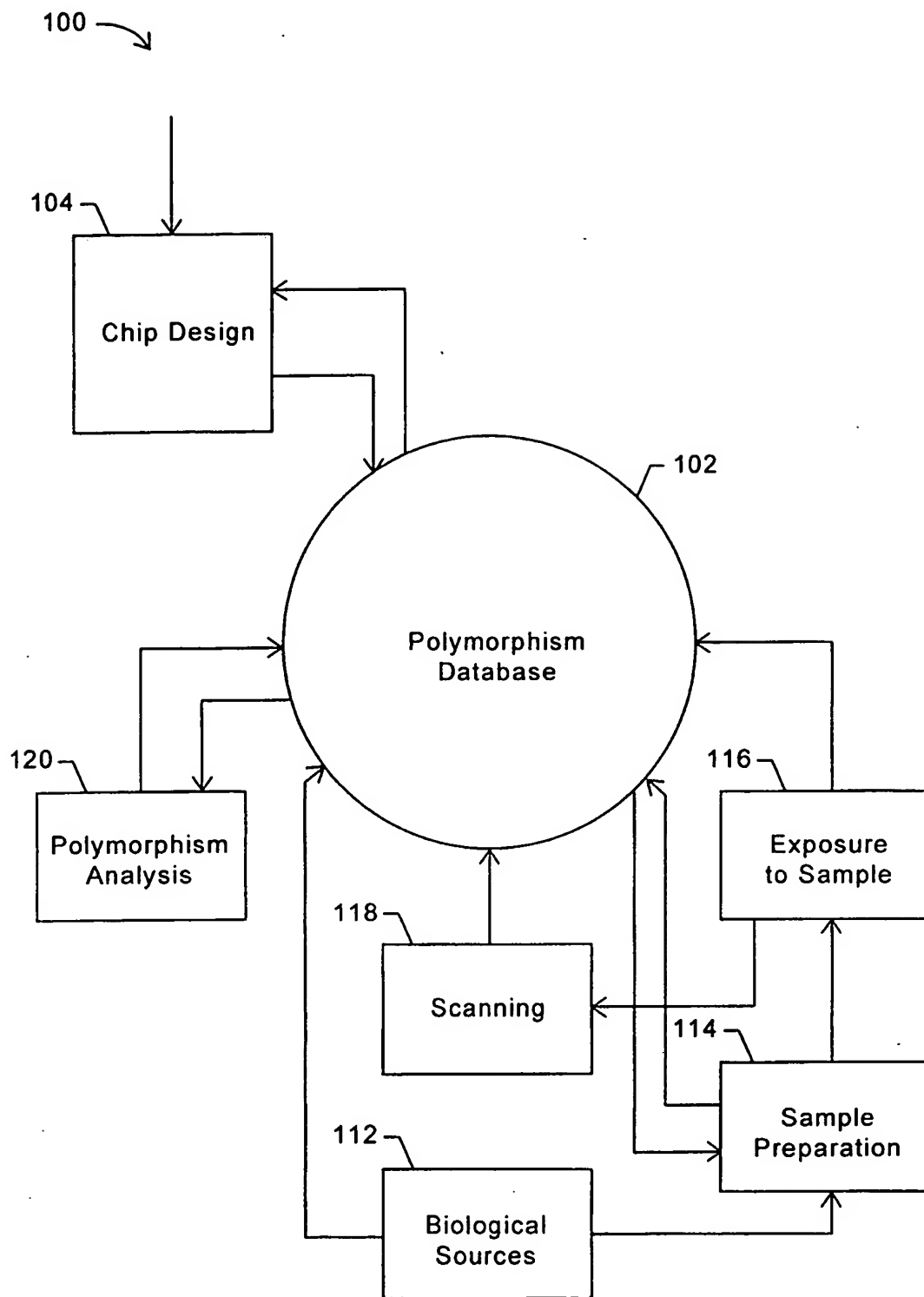


Fig. 1

2 / 12

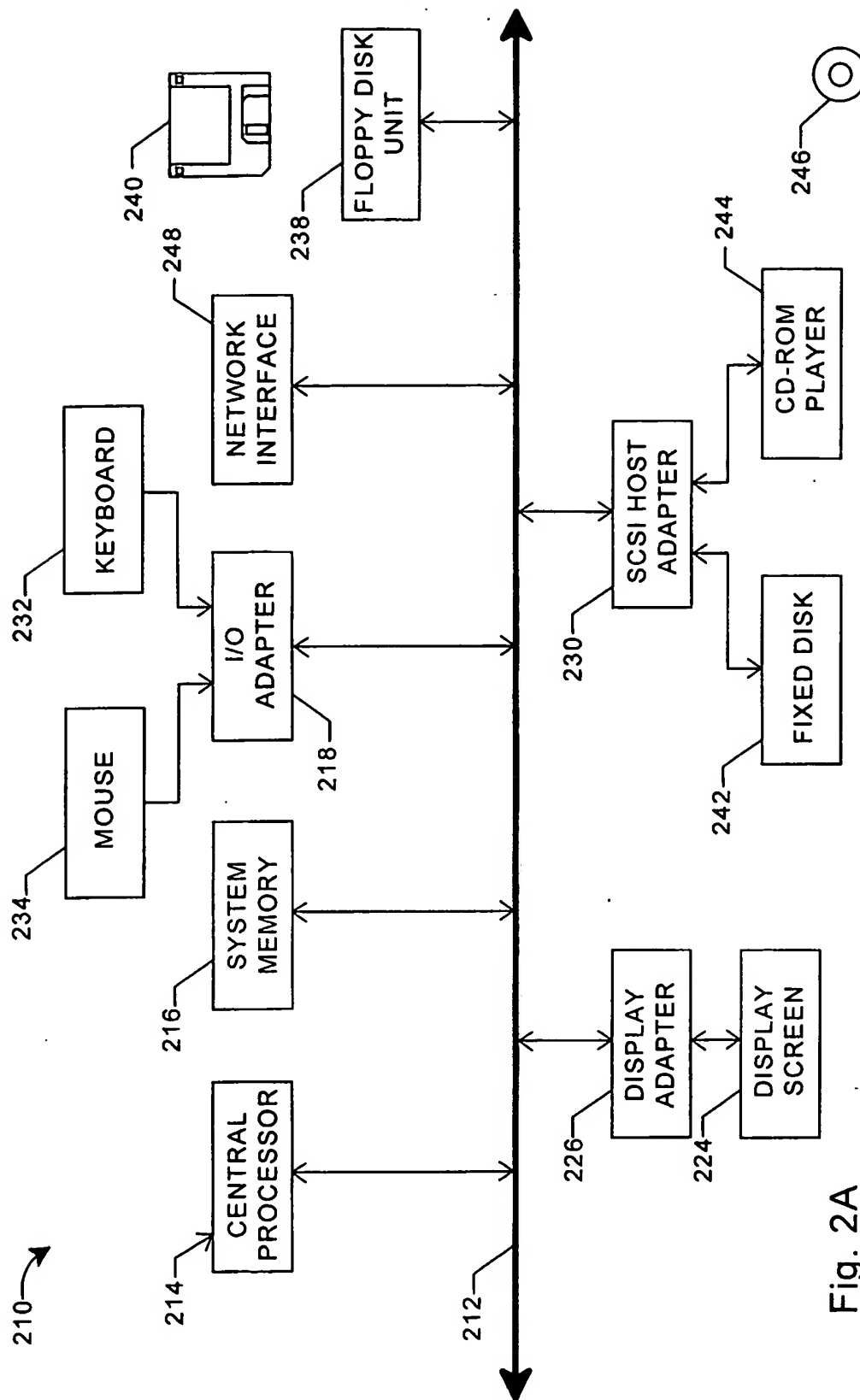


Fig. 2A

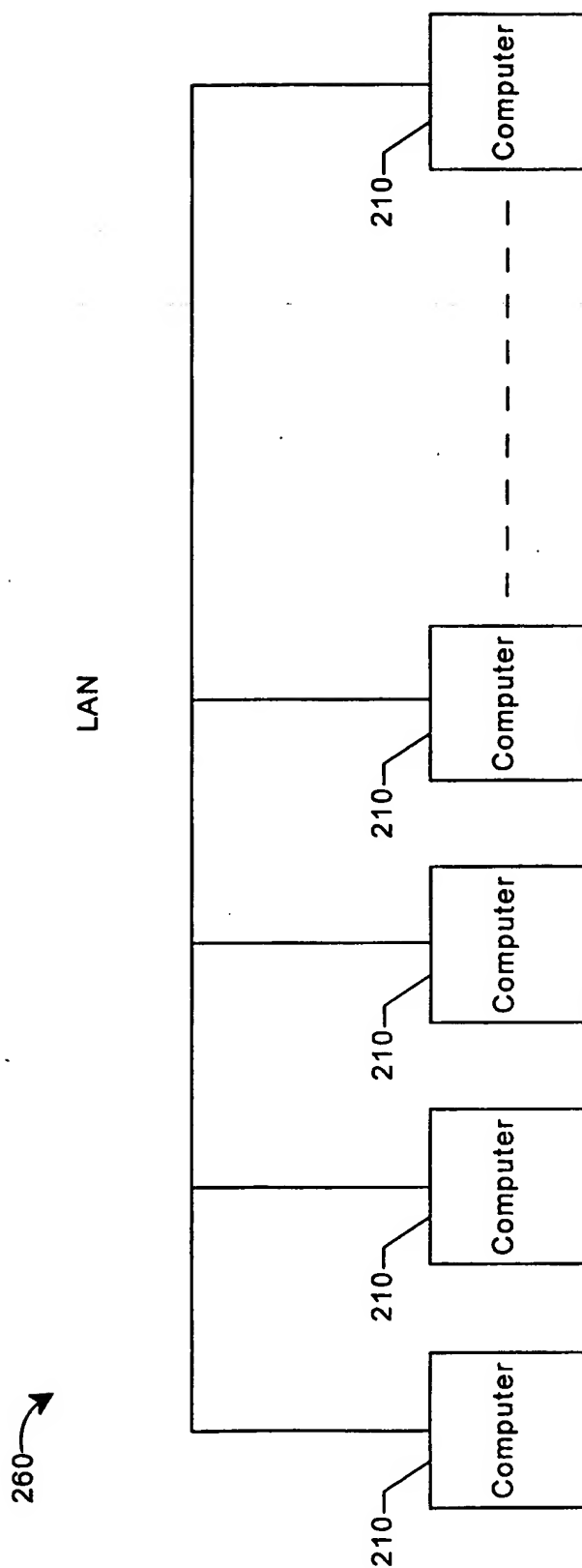


Fig. 2B

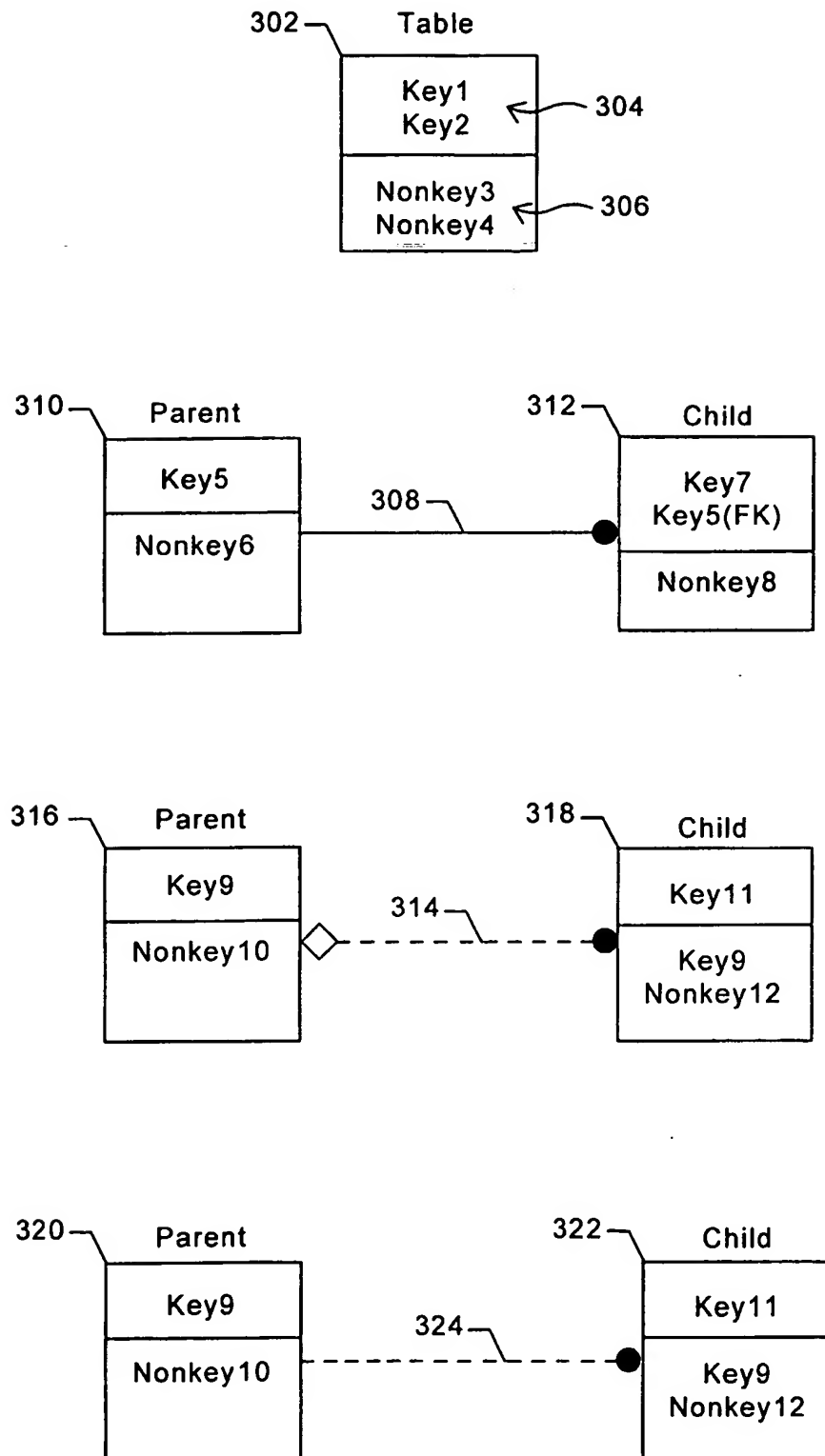


Fig. 3

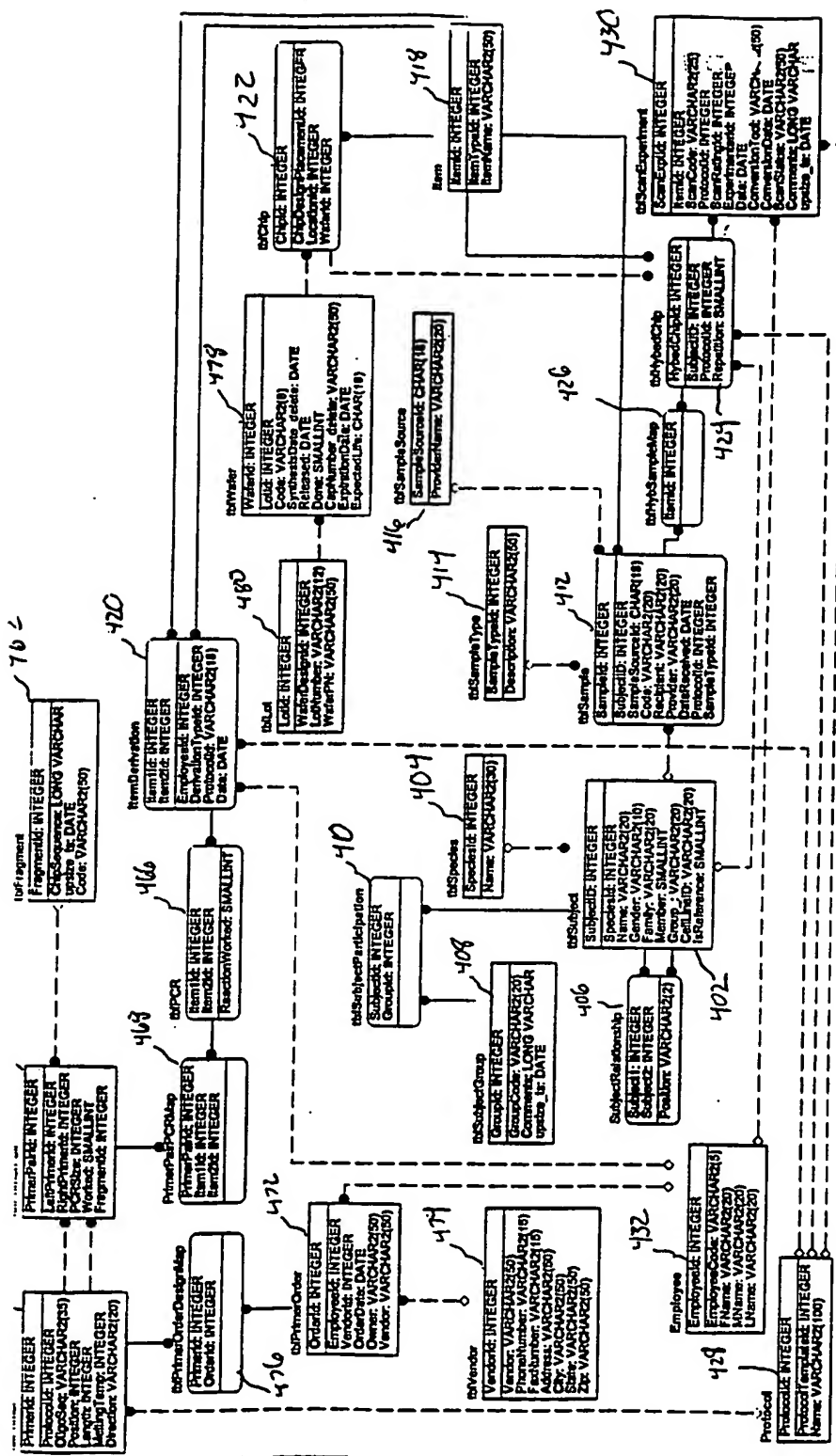


Fig. 4A

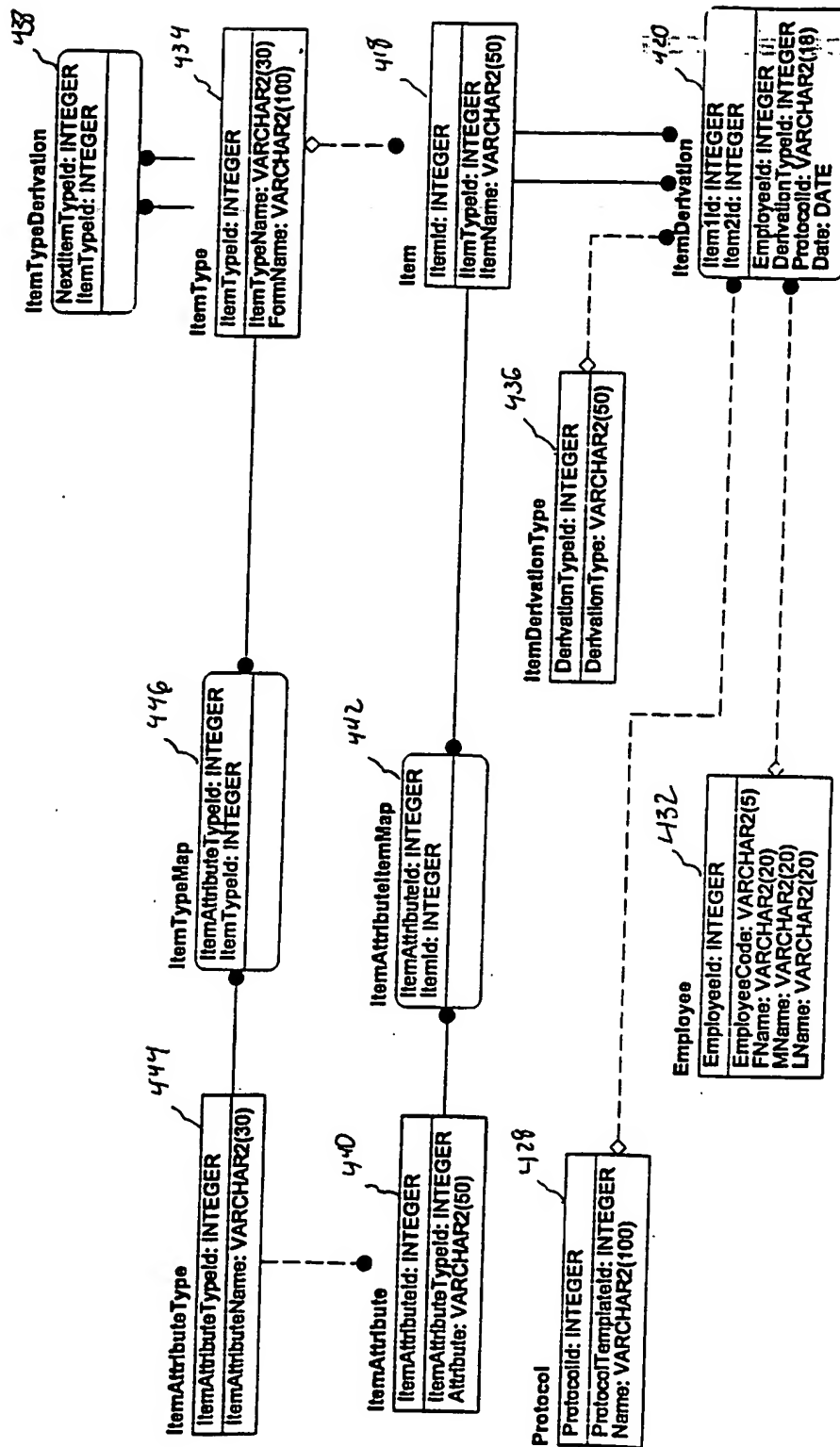


Fig. 4B

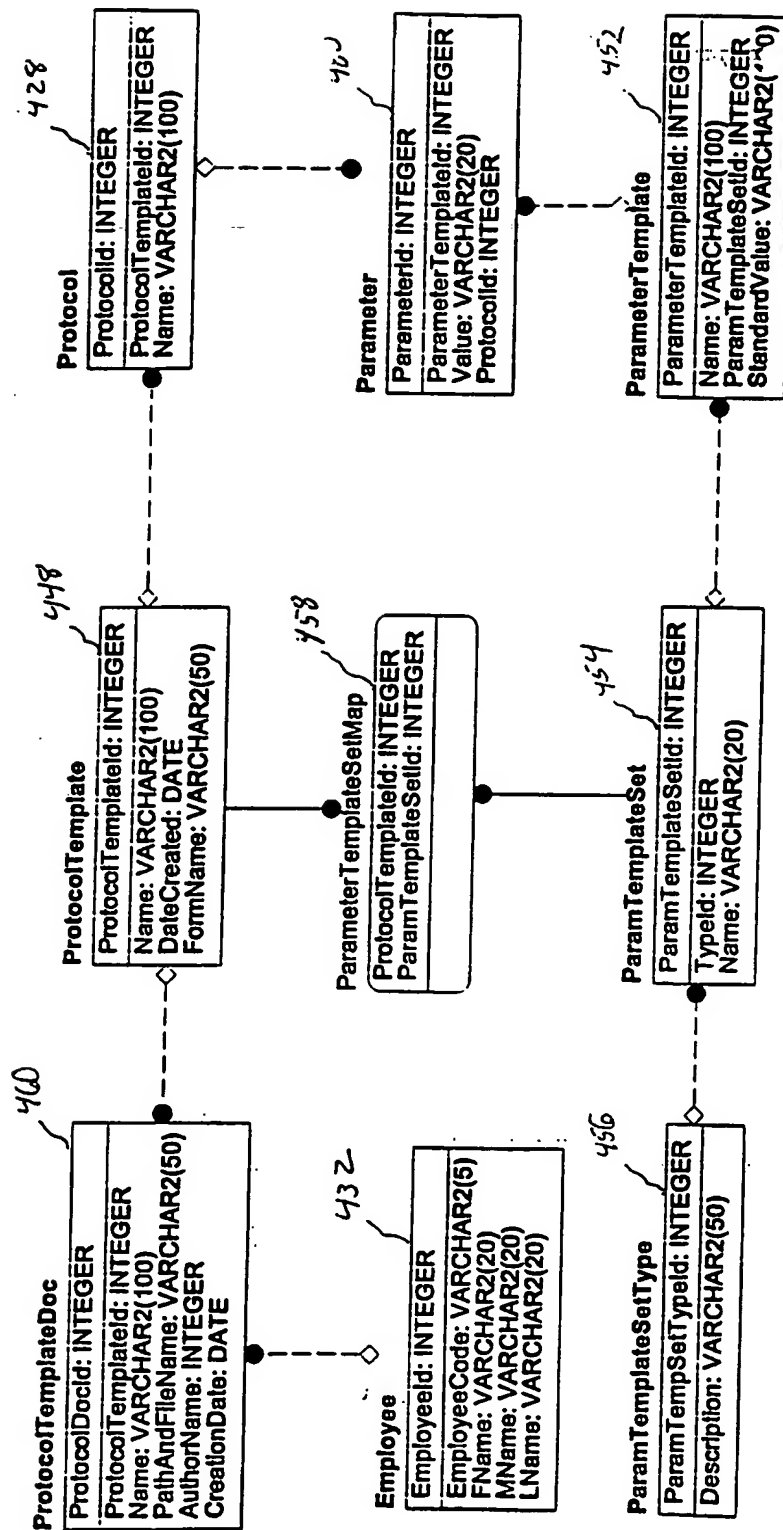


Fig. 4C



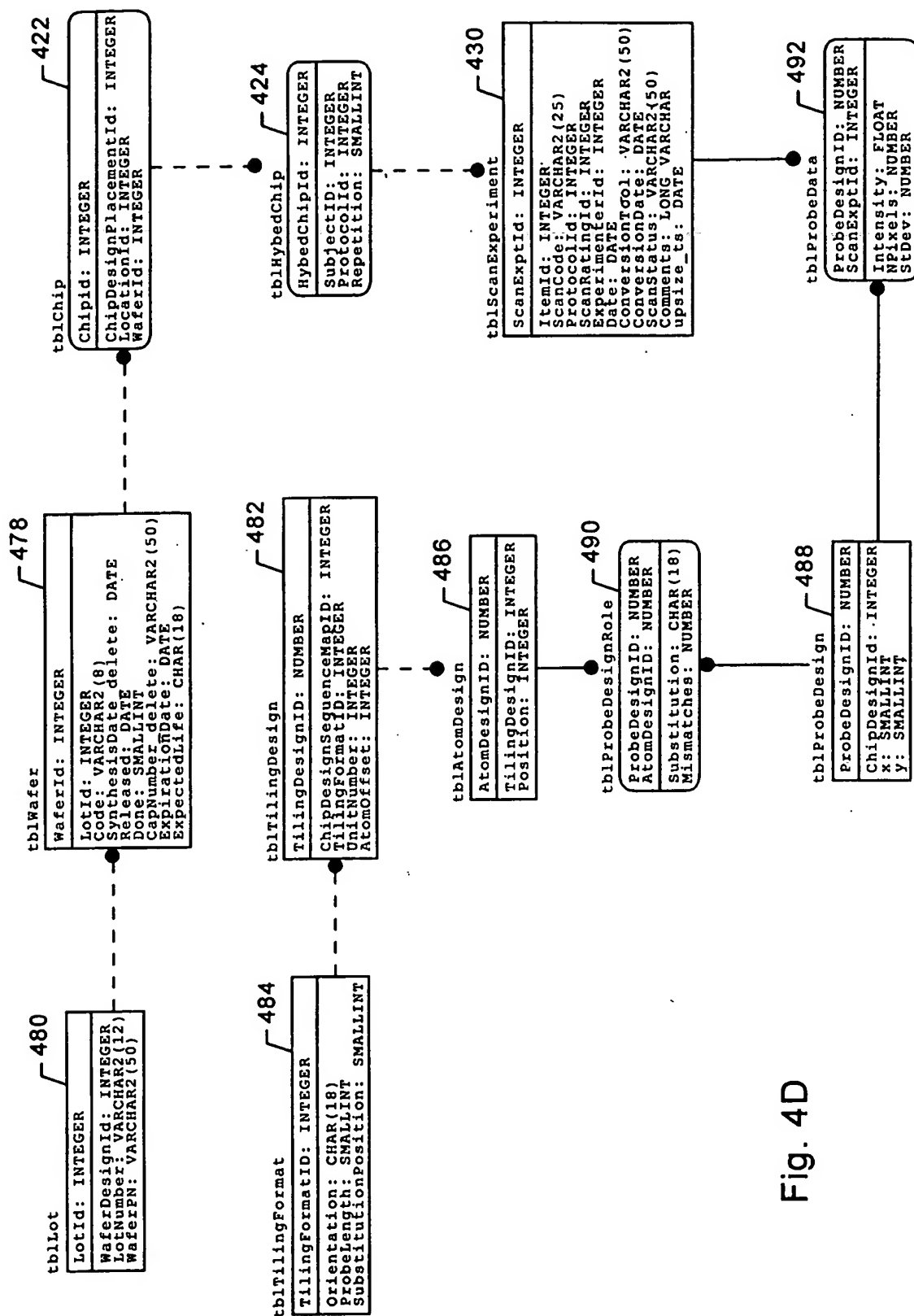


Fig. 4D

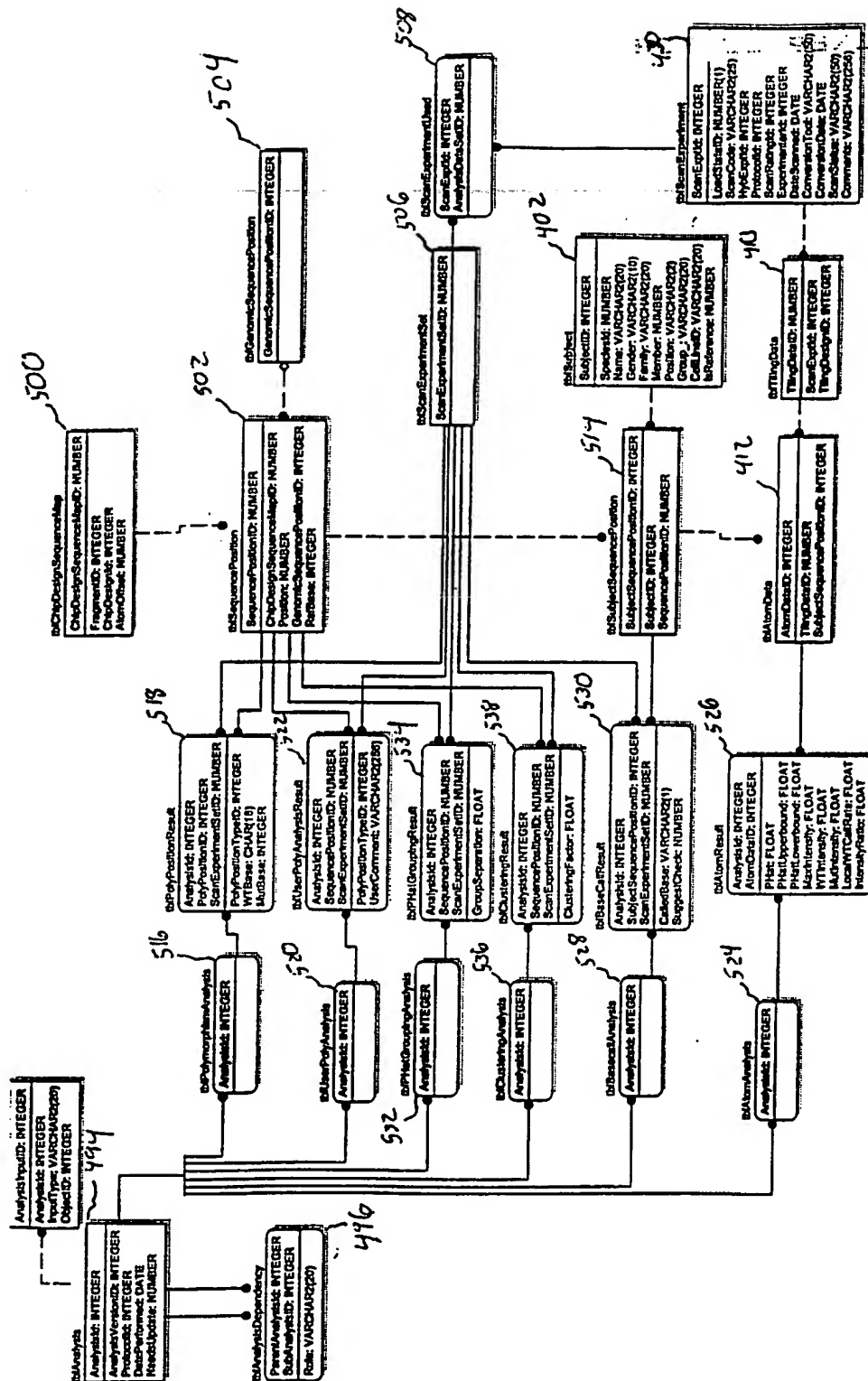


Fig. 25

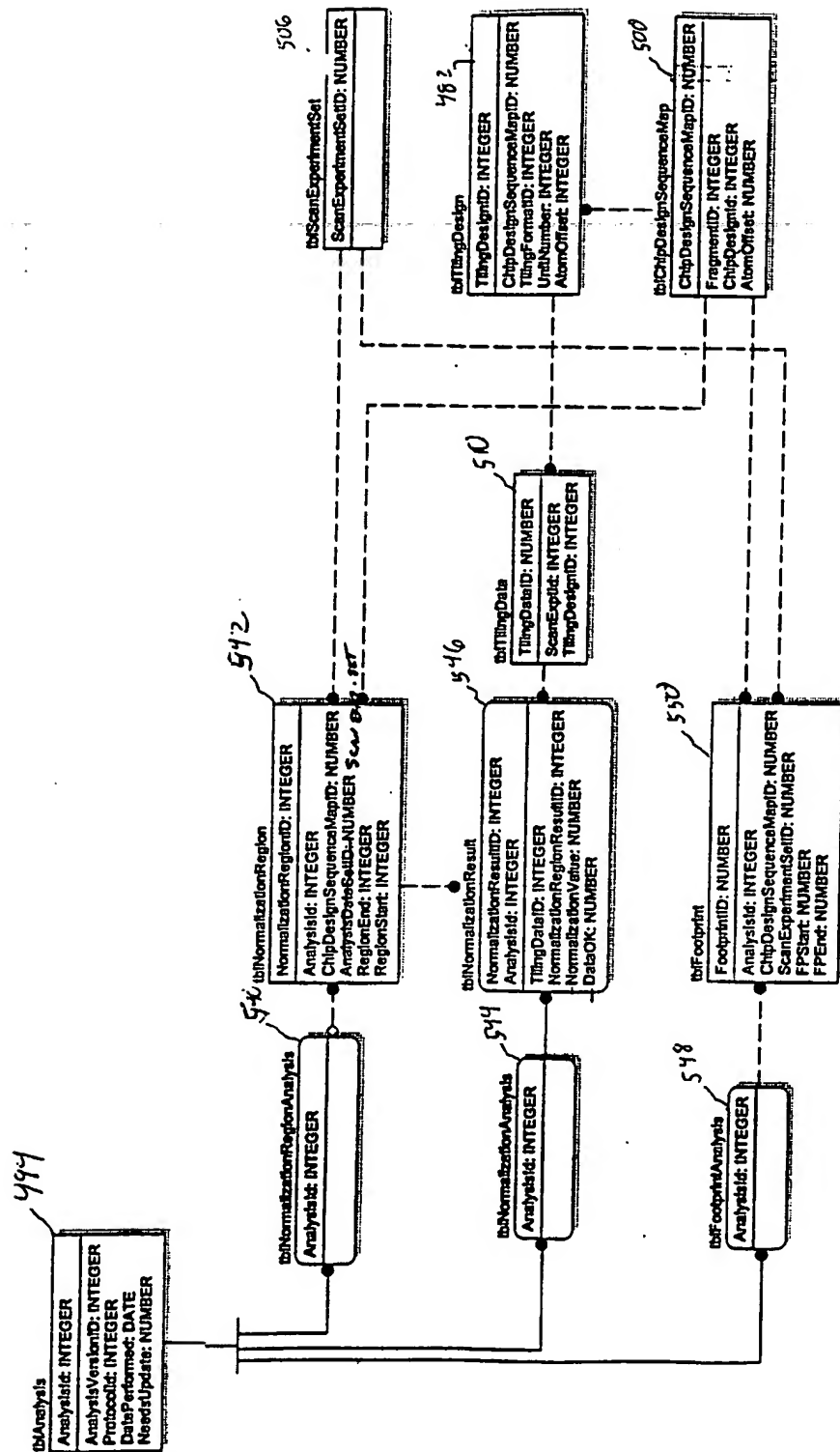


Fig. 4f

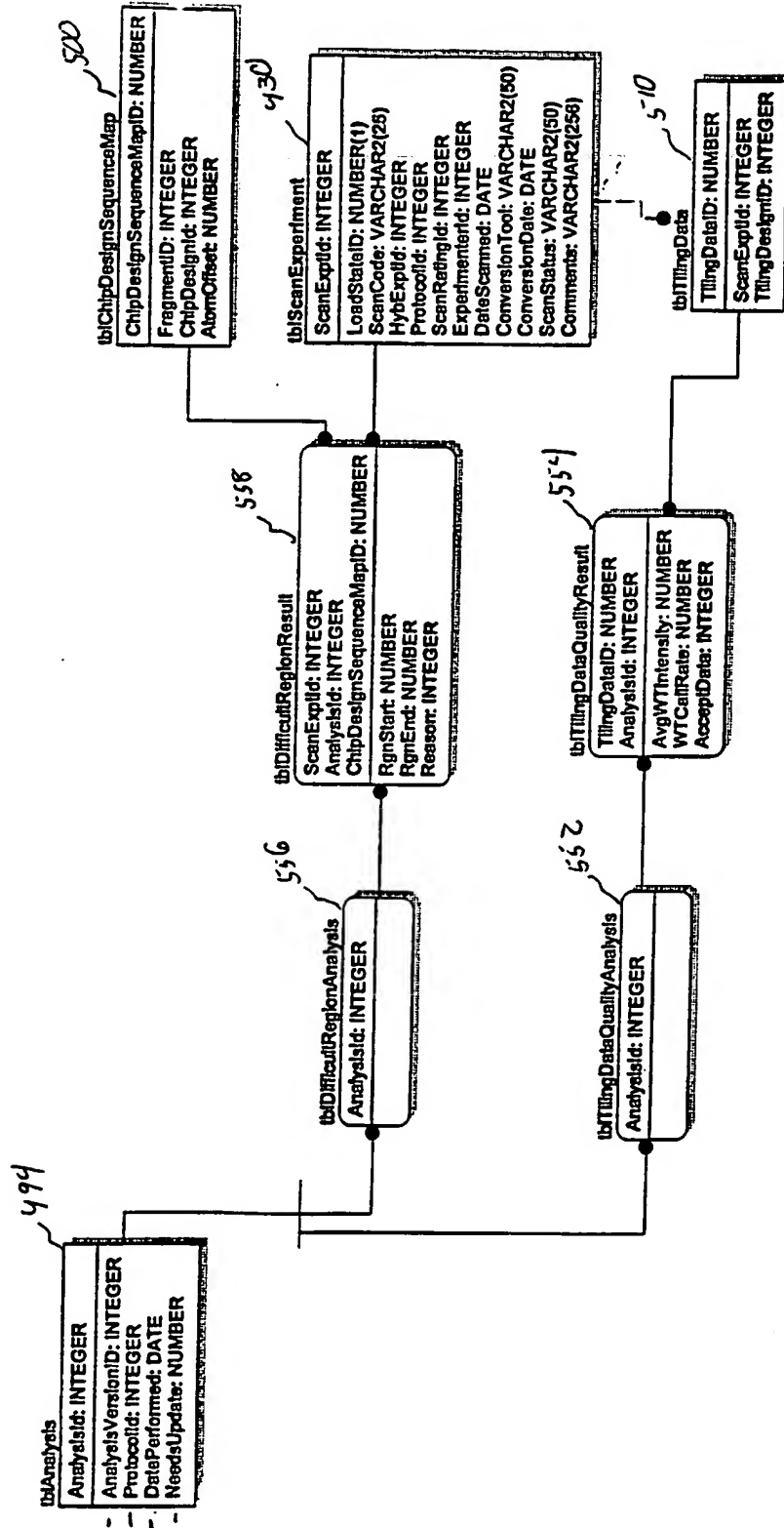


Fig. 46

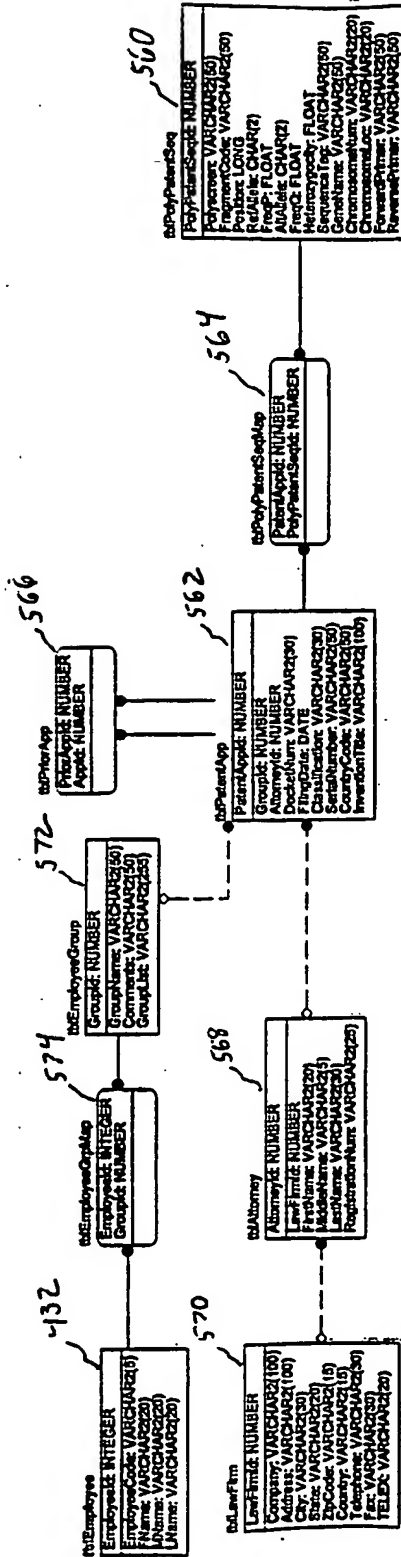


Fig. 7H

## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US98/15458**A. CLASSIFICATION OF SUBJECT MATTER**

IPC(6) : C12Q 1/68; G06F 17/30 // 159:00

US CL : 435/6; 707/104

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/6; 707/104

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

| Category* | Citation of document, with indication, where appropriate, of the relevant passages  | Relevant to claim No. |
|-----------|---|-----------------------|
| X         | WO 96/23078 A1 (INCITE PHARMACEUTICALS, INC) 01 AUGUST 1996, PAGE 1 LINE 26-PAGE 2, LINE 34 AND PAGE 5, LINE 17-PAGE 15, LINE 11. | 1-23                  |
| A         | US 5,206,137 A (IP ET AL) 27 APRIL 1993, COL. 5, LINE 25-COL. 6, LINE 9.  | 1-23                  |
| A         | US 5,593,839 A (HUBBELL ET AL) 14 JANUARY 1997, COL. 2, LINE 11-COL. 3, LINE 10.  | 1-23                  |
| AP        | US 5,707,806 A (SHUBER) 13 JANUARY 1998, COL. 2, LINES 37-68  | 1-23                  |

☐ Further documents are listed in the continuation of Box C.
 ☐ See patent family annex.

|   |  |
|---|--|
| * Special categories of cited documents:  | *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention  |
| *A* document defining the general state of the art which is not considered to be of particular relevance  | *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone   |
| *E* earlier document published on or after the international filing date  | *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art |
| *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) | *A* document member of the same patent family  |
| *O* document referring to an oral disclosure, use, exhibition or other means  |  |
| *P* document published prior to the international filing date but later than the priority date claimed  |  |

Date of the actual completion of the international search

01 OCTOBER 1998

Date of mailing of the international search report

29 OCT 1998

Name and mailing address of the ISA/US  
Commissioner of Patents and Trademarks  
Box PCT  
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

JACK M. CHOULES

Telephone No. (703) 305-9840

# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US98/15458

## B. FIELDS SEARCHED

Electronic data bases consulted (Name of data base and where practicable terms used):

APS, DIALOG, MAYA, NPL  
search terms DNA, cDNA, RNA, mRNA, gene, genome, genetic, data, database, relational, entity, relationship, many to many, polymorphism, database, microbiology, allele, mutant, sequence, analyze, heterozygosity.

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☒ FADED TEXT OR DRAWING
- ☒ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**